Access DB# _____

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # :_____ Date: _____
Art Unit: _____ Phone Number 30_____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

***For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.**

****

STAFF USE ONLY

Searcher: _____
Searcher Phone #: _____
Searcher Location: _____
Date Searcher Picked Up: _____
Date Completed: _____
Searcher Prep & Review Time: _____
Clerical Prep Time: _____
Online Time: _____

Type of Search

NA Sequence (#)_____
AA Sequence (#)_____
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis_____
Sequence Systems _____
WWW/Internet _____
Other (specify)_____

PTO-1590 (1-2000)

=> fil medline

FILE 'MEDLINE' ENTERED AT 12:57:13 ON 12 FEB 2001

FILE LAST UPDATED: 27 OCT 2000 (20001027/UP). FILE COVERS 1960 TO DATE.

MEDLINE thesauri in the /CN, /CT, and /MN fields incorporate the MeSH 2000 vocabulary. Enter HELP THESAURUS for details.

The OLDMEDLINE file segment now contains data from 1958 through 1965. Enter HELP CONTENT for details.

Left, right, and simultaneous left and right truncation are available in the Basic Index. See HELP SFIELDS for details.

THIS FILE CONTAINS CAS REGISTRY NUMBERS FOR EASY AND ACCURATE SUBSTANCE IDENTIFICATION.

MEDLINE UPDATES ARE ON HOLD UNTIL AFTER THE ANNUAL RELOAD HAS BEEN COMPLETED. NOTICE WILL BE GIVEN ONCE THE RELOAD IS COMPLETED AND RELOAD DETAILS WILL BE FOUND IN HELP RLOAD.

=> d all tot

L45 ANSWER 1 OF 30 MEDLINE
AN 2000354987 MEDLINE
DN 20311462
TI Computational genetics: finding protein function by nonhomology methods.
AU **Marcotte E M**
CS Molecular Biology Institute, UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California Los Angeles, Protein Pathways Inc., PO Box 951570, Los Angeles, Los Angeles, CA 90095-1570, CA 90024, USA.. marcotte@mbi.u
SO CURRENT OPINION IN STRUCTURAL BIOLOGY, (2000 Jun) 10 (3) 359-65.
Journal code: B9V. ISSN: 0959-440X.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200010
EW 20001002
AB During the past year, computational methods have been developed that use the rapidly accumulating genomic data to discover protein function. The methods rely on properties shared by functionally related proteins other than sequence or structural similarity. Instead, these 'nonhomology' methods analyze patterns such as domain fusion, conserved gene position and gene co-inheritance and coexpression to identify protein-protein relationships. The methods can identify functions for proteins that are without characterized homologs and have been applied to genome-wide predictions of protein function.
CT Check Tags: Animal; Human; Support, U.S. Gov't, Non-P.H.S.
Computer Simulation
Evolution, Molecular
Phylogeny
Proteins: AN, analysis
***Proteins: GE, genetics**
***Sequence Analysis: MT, methods**
Sequence Homology, Amino Acid
CN 6 (Proteins)

L45 ANSWER 2 OF 30 MEDLINE
AN 2000322577 MEDLINE
DN 20322577
TI Protein function in the post-genomic era.
AU **Eisenberg D; Marcotte E M**; Xenarios I; **Yeates T**

O

CS Molecular Biology Institute and UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California at Los Angeles, 90095-1570, USA.. david@mbi.ucla.edu

SO NATURE, (2000 Jun 15) 405 (6788) 823-6. Ref: 26
Journal code: NSC. ISSN: 0028-0836.

CY ENGLAND: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)

LA English

FS Priority Journals; Cancer Journals

EM 200009

EW 20000903

AB Faced with the avalanche of genomic sequences and data on messenger RNA expression, biological scientists are confronting a frightening prospect: piles of information but only flakes of knowledge. How can the thousands of sequences being determined and deposited, and the thousands of expression profiles being generated by the new array methods, be synthesized into useful knowledge? What form will this knowledge take? These are questions being addressed by scientists in the field known as 'functional genomics'.

CT Check Tags: Animal
Biotechnology
Computational Biology
Data Interpretation, Statistical
Genome
Phylogeny
Proteins: GE, genetics
***Proteins: PH, physiology**

CN 0 (Proteins)

L45 ANSWER 3 OF 30 MEDLINE

AN 2000183908 MEDLINE

DN 20183908

TI Selecting protein targets for structural genomics of Pyrobaculum aerophilum: validating automated fold assignment methods by using binary hypothesis testing.

AU Mallick P; Goodwill K E; Fitz-Gibbon S; Miller J H; **Eisenberg D**

CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Department of Chemistry and Biochemistry, Molecular Biology Institute, Box 951570, University of California, Los Angeles, CA 90095-1570, USA.

NC GM08375 (NIGMS)

SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (2000 Mar 14) 97 (6) 2450-5.
Journal code: PV3. ISSN: 0027-8424.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals; Cancer Journals

EM 200006

EW 20000605

AB Three-dimensional protein folds were assigned to all ORFs of the recently sequenced genome of the hyperthermophilic archaeon Pyrobaculum aerophilum. Binary hypothesis testing was used to estimate a confidence level for each assignment. A separate test was conducted to assign a probability for whether each sequence has a novel fold-i.e., one that is not yet represented in the experimental database of known structures. Of the 2,130 predicted nontransmembrane proteins in this organism, 916 matched a fold at a cumulative 90% confidence level, and 245 could be assigned at a 99% confidence level. Likewise, 286 proteins were predicted to have a previously unobserved fold with a 90% confidence level, and 14 at a 99% confidence level. These statistically based tools are combined with homology searches against the Online Mendelian Inheritance in Man (OMIM) human genetics database and other protein databases for the selection of attractive targets for crystallographic or NMR structure determination.

Results of these studies have been collated and placed at http://www.doe-mbi.ucla.edu/people/parag/P A_HOME/, the University of California, Los Angeles-Department of Energy Pyrobaculum aerophilum web site.
CT Check Tags: Human; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Algorithms
***Archaeal Proteins: CH, chemistry**
Computer Simulation
Databases, Factual
***Genome, Archaeal**
Membrane Proteins: CH, chemistry
Models, Chemical
Open Reading Frames
Protein Folding
Sequence Alignment: MT, methods
Thermoproteaceae: CH, chemistry
*Thermoproteaceae: GE, genetics
CN 0 (Archaeal Proteins); 0 (Membrane Proteins)

L45 ANSWER 4 OF 30 MEDLINE
AN 2000058633 MEDLINE
DN 20058633
TI Searching for frameshift evolutionary relationships between protein sequence families.
AU **Pellegrini M; Yeates T O**
CS Molecular Biology Institute, University of California, Los Angeles 90095-1570, USA.
SO PROTEINS, (1999 Nov 1) 37 (2) 278-83.
Journal code: PTS. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200003
EW 20000301
AB The protein sequence database was analyzed for evidence that some distinct sequence families might be distantly related in evolution by changes in frame of translation. Sequences were compared using special amino acid substitution matrices for the alternate frames of translation. The statistical significance of **alignment** scores were computed in the true database and shuffled versions of the database that preserve any potential codon bias. The comparison of results from these two databases provides a very sensitive method for detecting remote relationships. We find a weak but measurable relatedness within the database as a whole, supporting the notion that some proteins may have evolved from others through changes in frame of translation. We also quantify residual homology in the ordinary sense within a database of generally unrelated sequences.
CT Check Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.
Data Interpretation, Statistical
Databases, Factual
***Evolution, Molecular**
Frameshift Mutation
Point Mutation
***Proteins: CH, chemistry**
***Reading Frames**
Sequence Alignment
CN 0 (Proteins)

L45 ANSWER 5 OF 30 MEDLINE
AN 2000038092 MEDLINE
DN 20038092
TI A combined algorithm for genome-wide prediction of protein function [see comments].
CM Comment in: Nature 1999 Nov 4;402(6757):23, 25-6

AU **Marcotte E M; Pellegrini M; Thompson M J; Yeates T O; Eisenberg D**
CS Molecular Biology Institute, UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles 90095, USA.
SO NATURE, (1999 Nov 4) 402 (6757) 83-6.
Journal code: NSC. ISSN: 0028-0836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 200002
EW 20000204
AB The availability of over 20 fully sequenced genomes has driven the development of new methods to find protein function and interactions. Here we group proteins by correlated evolution, correlated messenger RNA expression patterns and patterns of domain fusion to determine functional relationships among the 6,217 proteins of the yeast Saccharomyces cerevisiae. Using these methods, we discover over 93,000 pairwise links between functionally related yeast proteins. Links between characterized and uncharacterized proteins allow a general function to be assigned to more than half of the 2,557 previously uncharacterized yeast proteins. Examples of functional links are given for a protein family of previously unknown function, a protein whose human homologues are implicated in colon cancer and the yeast prion Sup35.
CT Check Tags: Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.
***Algorithms**
Colorectal Neoplasms: ET, etiology
Evolution, Molecular
Fungal Proteins: CL, classification
Fungal Proteins: GE, genetics
***Fungal Proteins: PH, physiology**
Phylogeny
Prions: CL, classification
Prions: PH, physiology
RNA, Messenger: BI, biosynthesis
Saccharomyces cerevisiae
RN 133737-87-8 (SUP2 protein)
CN 0 (Fungal Proteins); 0 (MSH6 protein); 0 (Prions); 0 (RNA, Messenger)

L45 ANSWER 6 OF 30 MEDLINE
AN 1999443908 MEDLINE
DN 99443908
TI A census of protein repeats.
AU **Marcotte E M; Pellegrini M; Yeates T O; Eisenberg D**
CS Molecular Biology Institute, UCLA-DOE Lab of Structural Biology and Molecular Medicine, Los Angeles, CA, P.O. Box 951570, USA.
SO JOURNAL OF MOLECULAR BIOLOGY, (1999 Oct 15) 293 (1) 151-60.
Journal code: J6V. ISSN: 0022-2836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 200001
EW 20000104
AB In this study, we analyzed all known protein sequences for repeating amino acid segments. Although duplicated sequence segments occur in 14 % of all proteins, eukaryotic proteins are three times more likely to have internal repeats than prokaryotic proteins. After clustering the repetitive sequence segments into families, we find repeats from eukaryotic proteins have little similarity with prokaryotic repeats, suggesting most repeats arose after the prokaryotic and eukaryotic lineages diverged. Consequently, protein classes with the highest incidence of repetitive sequences perform functions unique to eukaryotes. The frequency distribution of the repeating units shows only weak length dependence,

implicating recombination rather than duplex melting or DNA hairpin formation as the limiting mechanism underlying repeat formation. The mechanism favors additional repeats once an initial duplication has been incorporated. Finally, we show that repetitive sequences are favored that contain small and relatively water-soluble residues. We propose that error-prone repeat expansion allows repetitive proteins to evolve more quickly than non-repeat-containing proteins. Copyright 1998 Academic Press.

CT Check Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.
Amino Acids: CH, chemistry
Databases
Eukaryotic Cells: CH, chemistry
Evolution, Molecular
Prokaryotic Cells: CH, chemistry
***Proteins: CH, chemistry**
Proteins: GE, genetics
***Repetitive Sequences, Nucleic Acid: GE, genetics**
Sequence Alignment
Solubility

CN 0 (Amino Acids); 0 (Proteins)

L45 ANSWER 7 OF 30 MEDLINE
AN 1999357871 MEDLINE
DN 99357871
TI Detecting protein function and protein-protein interactions from genome sequences.
AU **Marcotte E M; Pellegrini M;** Ng H L; Rice D W; **Yeates T O; Eisenberg D**
CS UCLA-Department of Energy Laboratory of Structural Biology and Molecular Medicine, University of California at Los Angeles, Los Angeles, CA 90095-1570, USA.
NC P01 GM 31299 (NIGMS)
SO SCIENCE, (1999 Jul 30) 285 (5428) 751-3.
Journal code: UJ7. ISSN: 0036-8075.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199910
EW 19991003
AB A computational method is proposed for inferring protein interactions from genome sequences on the basis of the observation that some pairs of interacting proteins have homologs in another organism fused into a single protein chain. Searching sequences from many genomes revealed 6809 such putative protein-protein interactions in Escherichia coli and 45,502 in yeast. Many members of these pairs were confirmed as functionally related; computational filtering further enriches for interactions. Some proteins have links to several other proteins; these coupled links appear to represent functional interactions such as complexes or pathways. Experimentally confirmed interacting pairs are documented in a Database of Interacting Proteins.
CT Check Tags: Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Amino Acid Sequence
Bacterial Proteins: CH, chemistry
Bacterial Proteins: GE, genetics
Bacterial Proteins: ME, metabolism
Bacterial Proteins: PH, physiology
Binding Sites
***Computational Biology**
Databases, Factual
Escherichia coli: GE, genetics
Evolution, Molecular
Fungal Proteins: CH, chemistry
Fungal Proteins: GE, genetics
Fungal Proteins: ME, metabolism

***Genome**
Genome, Bacterial
Genome, Fungal
Models, Biological
Proteins: CH, chemistry
Proteins: GE, genetics
Proteins: ME, metabolism
***Proteins: PH, physiology**
***Sequence Homology, Amino Acid**
***Sequence Homology, Nucleic Acid**
Thermodynamics
CN 0 (Bacterial Proteins); 0 (Fungal Proteins); 0 (Proteins)

L45 ANSWER 8 OF 30 MEDLINE
AN 1999321998 MEDLINE
DN 99321998
TI Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability [published erratum appears in J Mol Biol 1999 Oct 1;292(4):946].
AU **Thompson M J; Eisenberg D**
CS University of California Los Angeles, Los Angeles, CA 90095-1570, USA.
SO JOURNAL OF MOLECULAR BIOLOGY, (1999 Jul 9) 290 (2) 595-604.
Journal code: J6V. ISSN: 0022-2836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Cancer Journals; Priority Journals
EM 199910
AB Understanding the molecular determinants of protein thermostability is of theoretical and practical importance. While numerous determinants have been suggested, no molecular feature has been judged of paramount importance, with the possible exception of ion-pair networks. The difficulty in identifying the main determinants may have been the limited structural information available on the thermostable proteins. Recently the complete genomes for mesophilic, thermophilic and hyperthermophilic organisms have been sequenced, vastly improving the potential for uncovering general trends in sequence and structure evolution related to thermostability and, thus, for isolating the more important determinants. From a comparative analysis of 20 complete genomes, we find a trend towards shortened thermophilic proteins relative to their mesophilic homologs. Moreover, sequence **alignments** to proteins of known structure indicate that thermophilic sequences are more likely than their mesophilic homologs to have deletions in exposed loop regions. The new genomes offer enough comparable sequences to compute meaningful statistics that point to loop deletion as a general evolutionary strategy for increasing thermostability. Copyright 1999 Academic Press.
CT Check Tags: Comparative Study; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Archaeal Proteins: CH, chemistry
Archaeal Proteins: GE, genetics
Bacterial Proteins: CH, chemistry
Bacterial Proteins: GE, genetics
Databases, Factual
Enzyme Stability
***Evolution, Molecular**
Genome
Molecular Weight
Open Reading Frames
Phylogeny
Protein Conformation
***Proteins: CH, chemistry**
***Proteins: GE, genetics**
Sequence Alignment
*Sequence Deletion
Statistics
Temperature

Thermodynamics
CN 0 (Archaeal Proteins); 0 (Bacterial Proteins); 0 (Proteins)

L45 ANSWER 9 OF 30 MEDLINE
AN 1999310044 MEDLINE
DN 99310044
TI A fast algorithm for genome-wide analysis of proteins with repeated sequences.
AU **Pellegrini M; Marcotte E M; Yeates T O**
CS Molecular Biology Institute and UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, 90095-1570, U.S.A.
NC GM51299 (NIGMS)
SO PROTEINS, (1999 Jun 1; 35 (4) 440-6.
Journal code: PTS. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199910
EW 19991002
AB We present a fast algorithm to search for repeating fragments within protein sequences. The technique is based on an extension of the Smith-Waterman algorithm that allows the calculation of sub-optimal **alignments** of a sequence against itself. We are able to estimate the statistical significance of all sub-optimal **alignment** scores. We also rapidly determine the length of the repeating fragment and the number of times it is found in a sequence. The technique is applied to sequences in the Swissprot database, and to 16 complete genomes. We find that eukaryotic proteins contain more internal repeats than those of prokaryotic and archael organisms. The finding that 11% of yeast sequences and 28% of the known human sequences contain detectable repeats emphasizes the importance of internal duplication in protein evolution.
CT Check Tags: Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
***Algorithms**
Archaeoglobus fulgidus: GE, genetics
Databases, Factual
Escherichia coli: GE, genetics
Genes, Archaeal
Genome, Bacterial
Genome, Fungal
Poisson Distribution
***Proteins: CH, chemistry**
Saccharomyces cerevisiae: GE, genetics
Sequence Homology, Amino Acid
CN 0 (Proteins)

L45 ANSWER 10 OF 30 MEDLINE
AN 1999257335 MEDLINE
DN 99257335
TI Predicting structures for genome proteins.
AU Fischer D; **Eisenberg D**
CS Faculty of Natural Science, Department of Math and Computer Science, Beer-Sheva, 84015, Israel.. dfischer@cs.bgu.ac.il
SO CURRENT OPINION IN STRUCTURAL BIOLOGY, (1999 Apr) 9 (2) 208-11. Ref: 22
Journal code: B9V. ISSN: 0959-440X.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals
EM 199910
AB Assigning three-dimensional protein folds to genome sequences is essential to understanding protein function. Although experimental three-dimensional

structures are currently available for only a very small fraction of these sequences, computational fold assignment is able to assign folds to 20-30% of the sequences in various genomes. This percentage varies depending on the particular organism under analysis, on the sensitivities of the methods used and on the number of experimental structures available at the time the assignment is carried out. The fraction of assignable sequences is currently increasing at an annual rate of roughly 18%. If this rate is sustained throughout the coming years, three-dimensional computational models for more than half of the genome sequences may be available by the year 2003.

CT Check Tags: Human.
Amino Acid Sequence
Crystallization
Genome, Human
Internet
Models, Molecular
***Protein Conformation**
Protein Folding
***Proteins: CH, chemistry**
***Proteins: GE, genetics**
Sequence Homology, Amino Acid
CN 0 (Proteins)

L45 ANSWER 11 OF 30 MEDLINE
AN 1999218272 MEDLINE
DN 99218272
TI Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.
AU **Pellegrini M; Marcotte E M; Thompson M J; Eisenberg D; Yeates T O**
CS Molecular Biology Institute and Departments of Energy Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles, CA 90095-1570, USA.
SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1999 Apr 13) 96 (8) 4285-8.
Journal code: PV3. ISSN: 0027-8424.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199907
EW 19990704
AB Determining protein functions from genomic sequences is a central goal of bioinformatics. We present a method based on the assumption that proteins that function together in a pathway or structural complex are likely to evolve in a correlated fashion. During evolution, all such functionally linked proteins tend to be either preserved or eliminated in a new species. We describe this property of correlated evolution by characterizing each protein by its phylogenetic profile, a string that encodes the presence or absence of a protein in every known genome. We show that proteins having matching or similar profiles strongly tend to be functionally linked. This method of phylogenetic profiling allows us to predict the function of uncharacterized proteins.
CT Check Tags: Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Bacterial Proteins: CH, chemistry
Bacterial Proteins: GE, genetics
*Escherichia coli: GE, genetics
*Evolution, Molecular
***Genome**
***Genome, Bacterial**
Models, Biological
Open Reading Frames
*Phylogeny
***Proteins: CH, chemistry**
Proteins: GE, genetics

Ribosomal Proteins: CH, chemistry
CN 0 (ribosomal protein L7); 0 (Bacterial Proteins); 0 (Proteins); 0 (Ribosomal Proteins)

L45 ANSWER 12 OF 30 MEDLINE
AN 1999146450 MEDLINE
DN 99146450
TI Fold assignments for amino acid sequences of the CASP2 experiment.
AU Rice D W; Fischer D; Weiss R; **Eisenberg D**
CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine 90095-1570, USA.. dwrice@mbi.ucla.edu
SO PROTEINS, (1997) Suppl 1 11-22.
Journal code: PTS. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199805
EW 19980504
AB New and newly extended methods for fold assignment were tested for their abilities to assign folds to amino acid target sequences of unknown three-dimensional structure. These target sequences, released through the CASP2 experiment, are not obviously related to any sequence of known three-dimensional (3D) structure. We assigned 3D folds to target sequences and filed these predictions with CASP2 before their 3D structures were released. The methods tested were (1) Environmental 3D profiles of Bowie and colleagues [Bowie, J.U., Luthy, R., Eisenberg, D. Science 253:164-170, 1991]; (2) A variation of this is termed Directional Profiles; (3) The HSP2 five-dimensional sequence-structure substitution matrix of Rice and Eisenberg [Rice, D., Eisenberg, D.J. Mol. Biol. 267:1026-1037, 1997]; and (4) The Sequence Derived Property methods of Fischer and Eisenberg [Fischer, D., Eisenberg, D. Prot. Sci. 5:947-955, 1996]. When the 3D structures of the sequences were released, 17 of our predictions were evaluated. Of these 17, we assigned high probabilities to 11, of which 9 were correct. Five of these correct predictions were of known 3D structures similar to the targets and four of these were of new folds. The evaluation demonstrated that our methods were effective in assigning the proper fold to more than half of the CASP2 targets with known folds (5/9) and also were able to detect half of the sequences that corresponded to no known folds (4/8). Even when the correct fold is assigned to a sequence, proper **alignment** of the sequence to the structure remains a challenge. Our methods were able to produce accurate **alignments** (< 1.5 mean residue shift error from the structural **alignment**) for four of the targets, including the particularly difficult **alignment** (only 7% residue identity in the structurally **aligned** regions) of the ferrochelatase sequence to the fold of a periplasmic binding protein.
CT Check Tags: Support, U.S. Gov't, Non-P.H.S.
Amino Acid Sequence
Ferrochelatase: CH, chemistry
Molecular Sequence Data
***Protein Folding**
***Proteins: CH, chemistry**
Sequence Alignment
***Sequence Analysis**
CN EC 4.99.1.1 (Ferrochelatase); 0 (Proteins)

L45 ANSWER 13 OF 30 MEDLINE
AN 1998051679 MEDLINE
DN 98051679
TI Assessing the performance of fold recognition methods by means of a comprehensive benchmark.
AU Fischer D; Elofsson A; Rice D; **Eisenberg D**
CS UCLA-DOE Lab. of Structural Biology & Molecular Medicine, Molecular Biology Institute 90095-1570, USA.
SO PACIFIC SYMPOSIUM ON BIOCOMPUTING, (1996) 300-18.

Journal code: CWQ.
CY Singapore
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199803
EW 19980302
AB Recently there has been an explosion of methods for fold recognition. These methods seek to **align** a protein sequence to a three-dimensional structure and measure the compatibility of the sequence to the structure. In this work, we present a benchmark to assess the performance of such methods. The benchmark consists of a set of protein sequences matched by superposition to known structures. This set covers a wide range of protein families, and includes matching proteins with insignificant sequence similarity. To demonstrate the usefulness of this benchmark, we apply it here to compare different fold-recognition methods developed through the years in our group as well as several sequence-sequence substitution matrices. The results show that "global-local" **alignments** are superior to either local or global **alignments**. The most effective sequence-sequence matching matrix is the Gonnet table. The best performance overall is obtained by a method which combines the 3D-1D profiles of Bowie et al. with a substitution matrix and takes into account residue pairwise interactions.
CT Check Tags: Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.
***Amino Acid Sequence**
Computer Simulation
Enzymes: CH, chemistry
Molecular Biology: MT, methods
***Protein Folding**
***Protein Structure, Secondary**
***Proteins: CH, chemistry**
Sensitivity and Specificity
Sequence Alignment
Software
CN 0 (Enzymes); 0 (Proteins)

L45 ANSWER 14 OF 30 MEDLINE
AN 1998018099 MEDLINE
DN 98018099
TI VERIFY3D: assessment of protein models with three-dimensional profiles.
AU **Eisenberg D**; Luthy R; Bowie J U
CS Laboratory of Structural Biology and Molecular Medicine, University of California, Los Angeles 90095, USA.
SO METHODS IN ENZYMOLOGY, (1997) 277 396-404.
Journal code: MVA. ISSN: 0076-6879.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199801
EW 19980104
CT Check Tags: Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Amino Acid Sequence
Computer Communication Networks
***Computer Simulation**
Crystallography, X-Ray: MT, methods
Hemerythrin: CH, chemistry
Immunoglobulin Variable Region: CH, chemistry
***Models, Structural**
***Protein Conformation**
Protein Folding
***Proteins: CH, chemistry**
Reproducibility of Results
Ribulose-Bisphosphate Carboxylase: CH, chemistry
CN EC 4.1.1.39 (Ribulose-Bisphosphate Carboxylase); 0 (Hemerythrin); 0

(Immunoglobulin Variable Region); 0 (Proteins)

L45 ANSWER 15 OF 30 MEDLINE
AN 1998064500 MEDLINE
DN 98064500
TI Assigning folds to the proteins encoded by the genome of Mycoplasma genitalium.
AU Fischer D; **Eisenberg D**
CS University of California, Los Angeles-Department of Energy Laboratory of Structural Biology and Molecular Medicine, Molecular Biology Institute, University of California, Los Angeles, Box 951570, Los Angeles, CA 90095-1570, USA.
SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1997 Oct 28) 94 (22) 11929-34.
Journal code: PV3. ISSN: 0027-8424.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199802
EW 19980204
AB A crucial step in exploiting the information inherent in genome sequences is to assign to each protein sequence its three-dimensional fold and biological function. Here we describe fold assignment for the proteins encoded by the small genome of Mycoplasma genitalium. The assignment was carried out by our computer server (http://www.doe-mbi.ucla.edu/people/frsvr/ frsvr. html), which assigns folds to amino acid sequences by comparing sequence-derived predictions with known structures. Of the total of 468 protein ORFs, 103 (22%) can be assigned a known protein fold with high confidence, as cross-validated with tests on known structures. Of these sequences, 75 (16%) show enough sequence similarity to proteins of known structure that they can also be detected by traditional sequence-sequence comparison methods. That is, the difference of 28 sequences (6%) are assignable by the sequence-structure method of the server but not by current sequence-sequence methods. Of the remaining 78% of sequences in the genome, 18% belong to membrane proteins and the remaining 60% cannot be assigned either because these sequences correspond to no presently known fold or because of insensitivity of the method. At the current rate of determination of new folds by x-ray and NMR methods, extrapolation suggests that folds will be assigned to most soluble proteins in the next decade.
CT **Algorithms**
Amino Acid Sequence
***Bacterial Proteins: CH, chemistry**
Confidence Intervals
Databases, Factual
Forecasting
***Genome, Bacterial**
Hydrolases: CH, chemistry
Membrane Proteins: CH, chemistry
Models, Molecular
*Mycoplasma: GE, genetics
Nucleoside-Phosphate Kinase: CH, chemistry
***Protein Folding**
***Protein Structure, Secondary**
Sequence Alignment: MT, methods
Sequence Homology, Amino Acid
CN EC 2.7.4.4 (Nucleoside-Phosphate Kinase); EC 3. (Hydrolases); 0 (Bacterial Proteins); 0 (Membrane Proteins)

L45 ANSWER 16 OF 30 MEDLINE
AN 97445611 MEDLINE
DN 97445611
TI Predicting protein secondary structure with probabilistic schemata of evolutionarily derived information.
AU **Thompson M J**; Goldstein R A

CS Biophysics Research Division, University of Michigan, Ann Arbor 48109-1055, USA.
NC LM06770 (NLM)
SO PROTEIN SCIENCE, (1997 Sep) 6 (9) 1963-75.
Journal code: BNW. ISSN: 0961-8368.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199709
EW 19980305
AB We demonstrate the applicability of our previously developed Bayesian probabilistic approach for predicting residue solvent accessibility to the problem of predicting secondary structure. Using only single-sequence data, this method achieves a three-state accuracy of 67% over a database of 473 non-homologous proteins. This approach is more amenable to inspection and less likely to overlearn specifics for a dataset than "black box" methods such as neural networks. It is also conceptually simpler and less computationally costly. We also introduce a novel method for representing and incorporating multiple-sequence **alignment** information within the prediction algorithm, achieving 72% accuracy over a dataset of 304 non-homologous proteins. This is accomplished by creating a statistical model of the evolutionarily derived correlations between patterns of amino acid substitution and local protein structure. This model consists of parameter vectors, termed "substitution schemata," which probabilistically encode the structure-based heterogeneity in the distributions of amino acid substitutions found in **alignments** of homologous proteins. The model is optimized for structure prediction by maximizing the mutual information between the set of schemata and the database of secondary structures. Unlike "expert heuristic" methods, this approach has been demonstrated to work well over large datasets. Unlike the opaque neural network algorithms, this approach is physicochemically intelligible. Moreover, the model optimization procedure, the formalism for predicting one-dimensional structural features and our previously developed method for tertiary structure recognition all share a common Bayesian probabilistic basis. This consistency starkly contrasts with the hybrid and ad hoc nature of methods that have dominated this field in recent years.
CT Check Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Algorithms
Amino Acids: CH, chemistry
*Bayes Theorem
Chemistry, Physical
***Evolution, Molecular**
***Protein Structure, Secondary**
CN 0 (Amino Acids)

L45 ANSWER 17 OF 30 MEDLINE
AN 97280801 MEDLINE
DN 97280801
TI A 3D-1D substitution matrix for protein fold recognition that includes predicted secondary structure of the sequence.
AU Rice D W; **Eisenberg D**
CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Molecular Biology Institute, UCLA, Los Angeles, CA 90095-1570, USA.
NC GM07185 (NIGMS)
SO JOURNAL OF MOLECULAR BIOLOGY, (1997 Apr 11) 267 (4) 1026-38.
Journal code: JMV. ISSN: 0022-2836.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199707
EW 19970705
AB In protein fold recognition, a probe amino acid sequence is compared to a

library of representative folds of known structure to identify a structural homolog. In cases where the probe and its homolog have clear sequence similarity, traditional residue substitution matrices have been used to predict the structural similarity. In cases where the probe is sequentially distant from its homolog, we have developed a (7 x 3 x 2 x 7 x 3) 3D-1D substitution matrix (called H3P2), calculated from a database of 119 structural pairs. Members of each pair share a similar fold, but have sequence identity less than 30%. Each probe sequence position is defined by one of seven residue classes and three secondary structure classes. Each homologous fold position is defined by one of seven residue classes, three secondary structure classes, and two burial classes. Thus the matrix is five-dimensional and contains 7 x 3 x 2 x 7 x 3 = 882 elements or 3D-1D scores. The first step in assigning a probe sequence to its homologous fold is the prediction of the three-state (helix, strand, coil) secondary structure of the probe; here we use the profile based neural network prediction of secondary structure PHD program. Then a dynamic programming algorithm uses the H3P2 matrix to **align** the probe sequence with structures in a representative fold library. To test the effectiveness of the H3P2 matrix a challenging, fold class diverse, and cross-validated benchmark assessment is used to compare the H3P2 matrix to the GONNET, PAM250, BLOSUM62 and a secondary structure only substitution matrix. For distantly related sequences the H3P2 matrix detects more homologous structures at higher reliabilities than do these other substitution matrices, based on sensitivity versus specificity plots (or SENS-SPEC plots). The added efficacy of the H3P2 matrix arises from its information on the statistical preferences for various sequence-structure environment combinations from very distantly related proteins. It introduces the predicted secondary structure information from a sequence into fold recognition in a statistical way that normalizes the inherent correlations between residue type, secondary structure and solvent accessibility.

CT Check Tags: Comparative Study; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.

Algorithms
Amino Acids: CH, chemistry
***Databases, Factual**
Neural Networks (Computer)
***Protein Folding**
***Protein Structure, Secondary**
Proteins: CH, chemistry
***Sequence Alignment: MT, methods**
Sequence Homology, Amino Acid
Solvents

CN 0 (Amino Acids); 0 (Proteins); 0 (Solvents)

L45 ANSWER 18 OF 30 MEDLINE
AN 97235018 MEDLINE
DN 97235018
TI A study of combined structure/sequence profiles.
AU Elofsson A; Fischer D; Rice D W; Le Grand S M; **Eisenberg D**
CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, UCLA 90095-1570, USA.
NC 5-T32-GM07185 (NIGMS)
SO FOLDING AND DESIGN, (1996) 1 (6) 451-61.
Journal code: CUD. ISSN: 1359-0278.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199707
EW 19970705
AB BACKGROUND: For genome sequencing projects to achieve their full impact on biology and medicine, each protein sequence must be identified with its three-dimensional structure. Fold assignment methods (also called profile and threading methods) attempt to assign sequences to known protein folds by computing the compatibility of sequence to fold. RESULTS: We have

extended profile methods for the detection of protein folds having structural similarity but low sequence similarity to sequence probes. Our extension combines sequence substitution tables with structural properties to form a combined profile. The structural properties used in this study include distances between residues, exposed areas, areas buried by polar atoms, and properties of the original three-dimensional profile method. We compared the performance of these combined profiles with different sequence matrices and with the original three-dimensional profile method. To determine the optimal gap penalties and weights used with these profiles, we employed a genetic algorithm. The performance of these combined profiles was tested by cross validation using independent test and training sets. CONCLUSIONS: These studies show that the combined profiles perform better than profiles based on either structural or sequence information alone.

CT Check Tags: Animal; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Algorithms
Genome
Protein Conformation
***Proteins**
Proteins: CH, chemistry
Proteins: GE, genetics
Sequence Analysis

CN 0 (Proteins)

L45 ANSWER 19 OF 30 MEDLINE
AN 97185903 MEDLINE
IN 97185903
TI Into the black of night.
AU **Eisenberg D**
CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, 90025-1570, USA.. david@ewald.mbi.ucla.edu
SO NATURE STRUCTURAL BIOLOGY, (1997 Feb) 4 (2) 95-7.
Journal code: B98. ISSN: 1072-8368.
CY United States
DT Conference; Conference Article; (CONGRESSES)
LA English
FS Priority Journals
EM 199705
CT **Computer Communication Networks**
Computer Simulation
Evolution
Protein Conformation
***Protein Folding**
***Proteins: CH, chemistry**
Reproducibility of Results
Software
CN 0 (Proteins)

L45 ANSWER 20 OF 30 MEDLINE
AN 96338764 MEDLINE
IN 96338764
TI Three-dimensional profiles for measuring compatibility of amino acid sequence with three-dimensional structure.
AU Bowie J U; Zhang R; Wilmanns M; **Eisenberg D**
CS Department of Chemistry and Biochemistry, University of California, Los Angeles 90095, USA.
SO METHODS IN ENZYMOLOGY, (1996) 266 598-616.
Journal code: MVA. ISSN: 0076-6879.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199612
CT Check Tags: Comparative Study
Actins: CH, chemistry

*Amino Acid Sequence
Amino Acids: CH, chemistry
*Databases, Factual
Diphtheria Toxin: CH, chemistry
Heat-Shock Proteins 70: CH, chemistry
Mathematics
Models, Statistical
Models, Structural
Probability
*Protein Structure, Secondary
*Proteins: CH, chemistry
Reproducibility of Results
Sensitivity and Specificity
Software
CN 0 (Actins); 0 (Amino Acids); 0 (Diphtheria Toxin); 0 (Heat-Shock Proteins 70); 0 (Proteins)

L45 ANSWER 21 OF 30 MEDLINE
AN 96322957 MEDLINE
DN 96322957
TI Predicting solvent accessibility: higher accuracy using Bayesian statistics and optimized residue substitution classes.
AU **Thompson M J**; Goldstein R A
CS Biophysics Research Division, University of Michigan, Ann Arbor 48109-1055, USA.
NC R29 LM05770 (NLM)
SO PROTEINS, (1996 May) 25 (1) 38-47.
Journal code: PTS. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199701
EW 19970104
AB We introduce a novel Bayesian probabilistic method for predicting the solvent accessibilities of amino acid residues in globular proteins. Using single sequence data, this method achieves prediction accuracies higher than previously published methods. Substantially improved predictions-comparable to the highest accuracies reported in the literature to date-are obtained by representing **alignments** of the example proteins and their homologs as strings of residue substitution classes, depending on the side chain types observed at each **alignment** position. These results demonstrate the applicability of this relatively simple Bayesian approach to structure prediction and illustrate the utility of the classification methodology previously developed to extract information from **aligned** sets of structurally related proteins.
CT Check Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Amino Acid Sequence
***Amino Acids: CH, chemistry**
Bayes Theorem
Databases, Factual
Evolution, Molecular
Information Theory
Likelihood Functions
Molecular Sequence Data
Neural Networks (Computer)
***Protein Conformation**
Protein Folding
Protein Structure, Secondary
***Proteins: CH, chemistry**
Sequence Alignment
Solvents
CN 0 (Amino Acids); 0 (Proteins); 0 (Solvents)

L45 ANSWER 22 OF 30 MEDLINE

AN 96303956 MEDLINE
DN 96303956
TI Constructing amino acid residue substitution classes maximally indicative of local protein structure.
AU **Thompson M J**; Goldstein R A
CS Biophysics Research Division, University of Michigan, Ann Arbor 48109-1055, USA.
NC R29 LM05770 (NLM)
SO PROTEINS, (1996 May) 25 (1) 28-37.
Journal code: PTS. ISSN: 0887-3585.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199701
EW 19970104
AB Using an information theoretic formalism, we optimize classes of amino acid substitution to be maximally indicative of local protein structure. Our statistically-derived classes are loosely identifiable with the heuristic constructions found in previously published work. However, while these other methods provide a more rigid idealization of physicochemically constrained residue substitution, our classes provide substantially more structural information with many fewer parameters. Moreover, these substitution classes are consistent with the paradigmatic view of the sequence-to-structure relationship in globular proteins which holds that the three-dimensional architecture is predominantly determined by the arrangement of hydrophobic and polar side chains with weak constraints on the actual amino acid identities. More specific constraints are imposed on the placement of prolines, glycines, and the charged residues. These substitution classes have been used in highly accurate predictions of residue solvent accessibility. They could also be used in the identification of homologous proteins, the construction and refinement of multiple sequence **alignments**, and as a means of condensing and codifying the information in multiple sequence **alignments** for secondary structure prediction and tertiary fold recognition.
CT Check Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Amino Acid Sequence
***Amino Acids: CH, chemistry**
Databases, Factual
Information Theory
Molecular Sequence Data
***Protein Conformation**
Protein Structure, Secondary
Protein Structure, Tertiary
***Proteins: CH, chemistry**
Sequence Alignment
Solvents
CN 0 (Amino Acids); 0 (Proteins); 0 (Solvents)

L45 ANSWER 23 OF 30 MEDLINE
AN 96147301 MEDLINE
DN 96147301
TI Assigning amino acid sequences to 3-dimensional protein folds.
AU Fischer D; Rice D; Bowie J U; **Eisenberg D**
CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine, Molecular Biology Institute 90095-1570, USA.
SO FASEB JOURNAL, (1996 Jan) 10 (1) 126-36. Ref: 54
Journal code: FAS. ISSN: 0892-6638.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
General Review; (REVIEW)
(REVIEW, TUTORIAL)
LA English
FS Priority Journals; Cancer Journals
EM 199605
AB With the advent of genome sequencing projects, the amino acid sequences of

thousands of proteins are determined every year. Each of these protein sequences must be identified with its function and its 3-dimensional structure for us to gain a full understanding of the molecular biology of organisms. To meet this challenge, new methods are being developed for fold recognition, the computational assignment of newly determined amino acid sequences to 3-dimensional protein structures. These methods start with a library of known 3-dimensional target protein structures. The new probe sequence is then **aligned** to each target protein structure in the library and the compatibility of the sequence for that structure is scored. If a target structure is found to have a significantly high compatibility score, it is assumed that the probe sequence folds in much the same way as the target structure. The fundamental assumptions of this approach are that many different sequences fold in similar ways and there is a relatively high probability that a new sequence possesses a previously observed fold. We review various approaches to fold recognition and break down the process into its main steps: creation of a library of target folds; representation of the folds; **alignment** of the probe sequence to a target fold using a sequence-to-structure compatibility scoring function; and assessment of significance of compatibility. We emphasize that even though this new field of fold recognition has made rapid progress, technical problems remain to be solved in most of the steps. Standard benchmarks may help identify the problem steps and find solutions to the problems.

CT **Algorithms**
Computer Graphics
Databases, Factual
Forecasting
***Protein Conformation**
***Protein Folding**
Sequence Alignment
***Sequence Analysis: MT, methods**

L45 ANSWER 24 OF 30 MEDLINE
AN 94272351 MEDLINE
DN 94272351
TI The three-dimensional profile method using residue preference as a continuous function of residue environment.
AU Zhang K Y; **Eisenberg D**
CS UCLA-DOE Laboratory of Structural Biology and Molecular Medicine 90024-1570..
SO PROTEIN SCIENCE, (1994 Apr) 3 (4) 687-95.
Journal code: BNW. ISSN: 0961-8368.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199409
AB In the 3-dimensional profile method, the compatibility of an amino acid sequence for a given protein structure is scored as the sum of the preferences of the residues for their environments in the 3D structure. In the original method (Bowie JU, Luthy R, Eisenberg D, 1991, Science 253:164-170), residue environments were quantized into 18 discrete environmental classes. Here, amino acid residue preferences are expressed as a continuous function of environmental variables (residue area buried and fractional area buried by polar atoms). This continuous representation of residue preferences, expressed as a Fourier series, avoids the abrupt change of preference of residues in slightly different environments, as encountered in the original method with its 18 discrete environmental classes. When compared with the discrete 18-class representation of residue environments, this continuous 3D profile is found to be more sensitive in identifying sequences that fold into the profiled structure but share with it little sequence identity. The continuous 3D profile is also less sensitive to errors in environmental variables than is the discrete 3D profile. The continuous 3D profile can also be used to detect wrong folds or incorrectly modeled segments in an otherwise correct structure, as could the discrete 3D profile (Luthy R, Bowie JU, Eisenberg

D, 1992, Nature 356:83-85). Moreover, the progress of structure improvement during atomic refinement can also be monitored by examining the profile scores in a moving-window scan. Finally, by defining a functional form for profile scores, we open the way to profile atomic refinement in which an atomic structure adjusts to produce residue environments more compatible with the protein side chains.

CT Check Tags: Animal; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.

***Amino Acids: CH, chemistry**
Carp
Cattle
Dogfish
Lactate Dehydrogenase: CH, chemistry
Models, Chemical
Muscles: EN, enzymology
Myoglobin: CH, chemistry
Parvalbumins: CH, chemistry
***Protein Folding**
Protein Structure, Secondary
Sequence Alignment
Sequence Analysis
Sequence Homology, Amino Acid
Software
Superoxide Dismutase: CH, chemistry
Whales

CN EC 1.1.1.27 (Lactate Dehydrogenase); EC 1.15.1.1 (Superoxide Dismutase); 0 (Amino Acids); 0 (Myoglobin); 0 (Parvalbumins)

L45 ANSWER 25 OF 30 MEDLINE
AN 94240152 MEDLINE
DN 94240152
TI An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function.
AU Bowie J U; **Eisenberg D**
CS Department of Chemistry and Biochemistry, University of California, Los Angeles..
SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1994 May 10) 91 (10) 4436-40.
Journal code: PV3. ISSN: 0027-8424.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199408
AB Three short protein sequences have been guided by computer to folds resembling their crystal structures. Initially, peptide fragment conformations ranging in size from 9 to 25 residues were selected from a database of known protein structures. A fragment was selected if it was compatible with a segment of the sequence to be folded, as judged by three-dimensional profile scores. By linking the selected fragment conformations together, hundreds of trial structures were generated of the same length and sequence as the protein to be folded. These starting trial structures were then improved by an evolutionary algorithm. Selection pressure for improving the structures was provided by an energy function that was designed to guide the conformational search procedure toward the correct structure. We find that by evolution of only 400 structures for fewer than 1400 generations, the overall fold of some small helical proteins can be computed from the sequence, with deviations from observed structures of 2.5-4.0 A for C alpha atoms.

CT Check Tags: Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.

Algorithms
***Amino Acid Sequence**
***Evolution**
Models, Genetic
Models, Molecular

Mutation
***Protein Conformation**
***Protein Folding**
***Protein Structure, Secondary**
***Proteins: CH, chemistry**
Proteins: GE, genetics
Recombination, Genetic
Statistics
CN 0 (Proteins)

L45 ANSWER 26 OF 30 MEDLINE
AN 93165700 MEDLINE
DN 93165700
TI Three-dimensional profiles from residue-pair preferences: identification of sequences with beta/alpha-barrel fold.
AU Wilmanns M; **Eisenberg D**
CS Molecular Biology Institute, University of California, Los Angeles 90024-1570..
SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1993 Feb 15) 90 (4) 1379-83.
Journal code: PV3. ISSN: 0027-8424.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Cancer Journals; Priority Journals
EM 199305
AB The three-dimensional profile method expresses the three-dimensional structure of a protein as a table, the profile, which represents the local environment of each residue. The score of an amino acid sequence, **aligned** with the three-dimensional profile, reflects its compatibility with the profiled structure. In the original implementation, each local environment was characterized by its polarity, the area buried of its side chain, and its secondary structure. Here we describe a modified three-dimensional profile algorithm that characterizes the local environment in terms of the statistical preferences of the profiled residue for neighbors of specific residue types, main-chain conformations, or secondary structure. Combined profiles of the original and the three new types were tested on beta/alpha-barrel protein structures. The method identified the following enzymes of unknown three-dimensional structure as probable beta/alpha-barrels, all of which catalyze reactions in the biosynthesis of aromatic amino acids: anthranilate phosphoribosyltransferase (trpD), glutamine amidotransferase (trpG), and phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (hisA).
CT Check Tags: Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.
Algorithms
***Amino Acid Sequence**
Anthranilate Phosphoribosyltransferase: CH, chemistry
Databases, Factual
*Enzymes: CH, chemistry
Isomerases: CH, chemistry
*Models, Molecular
***Protein Conformation**
***Protein Folding**
Protein Structure, Secondary
***Proteins: CH, chemistry**
Transferases: CH, chemistry
CN EC 2. (Transferases); EC 2.4.2.18 (Anthranilate Phosphoribosyltransferase); EC 2.6.- (glutamine amidotransferase); EC 5. (Isomerases); EC 5.3.1.16 (N-(5'-phosphoribosylformimino)-5-amino-1-(5''-phosphoribosyl)-4-imidazolecarboxamide isomerase); C (Enzymes); 0 (Proteins)

L45 ANSWER 27 OF 30 MEDLINE
AN 91305934 MEDLINE

DN 91305934
TI A method to identify protein sequences that fold into a known three-dimensional structure.
AU Bowie J U; Luthy R; **Eisenberg D**
CS Molecular Biology Institute, University of California, Los Angeles 90024-1570..
SO SCIENCE, (1991 Jul 12) 253 (5016) 164-70.
Journal code: 0404. ISSN: 0036-8075.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 199110
AB The inverse protein folding problem, the problem of finding which amino acid sequences fold into a known three-dimensional (3D) structure, can be effectively attacked by finding sequences that are most compatible with the environments of the residues in the 3D structure. The environments are described by: (i) the area of the residue buried in the protein and inaccessible to solvent; (ii) the fraction of side-chain area that is covered by polar atoms (O and N); and (iii) the local secondary structure. Examples of this 3D profile method are presented for four families of proteins: the globins, cyclic AMP (adenosine 3',5'-monophosphate) receptor-like proteins, the periplasmic binding proteins, and the actins. This method is able to detect the structural similarity of the actins and 70- kilodalton heat shock proteins, even though these protein families share no detectable sequence similarity.
CT Check Tags: Animal; Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Actins: CH, chemistry
Actins: UL, ultrastructure
Algorithms
Amino Acid Sequence
Carrier Proteins: CH, chemistry
Molecular Structure
Myoglobin: CH, chemistry
Myoglobin: UL, ultrastructure
***Protein Conformation**
***Proteins: CH, chemistry**
Receptors, Cyclic AMP: CH, chemistry
Receptors, Cyclic AMP: UL, ultrastructure
Structure-Activity Relationship
CN 0 (ribose-binding protein); 0 (Actins); 0 (Carrier Proteins); 0 (Myoglobin); 0 (Receptors, Cyclic AMP)

L45 ANSWER 28 OF 30 MEDLINE
AN 90190364 MEDLINE
DN 90190364
TI Profile analysis.
AU Gribskov M; Luthy R; **Eisenberg D**
NC N01-CO-74101 (NCI)
SO METHODS IN ENZYMOLOGY, (1990) 183 146-59.
Journal code: MVA. ISSN: 0076-6879.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199006
CT Check Tags: Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
***Amino Acid Sequence**
***Information Systems**
Molecular Sequence Data
***Proteins: GE, genetics**
*Research: MT, methods
Sequence Homology, Nucleic Acid
***Software**

L45 ANSWER 29 OF 30 MEDLINE
AN 88052389 MEDLINE
DN 88052389
TI Profile scanning for three-dimensional structural patterns in protein sequences.
AU Gribskov M; Homyak M; Edenfield J; **Eisenberg D**
CS Molecular Biology Institute, University of California, Los Angeles 90024-1570.
NC CA-040096 (NCI)
SO COMPUTER APPLICATIONS IN THE BIOSCIENCES, (1988 Mar) 4 (1) 61-6.
Journal code: CAB. ISSN: 0266-7061.
CY ENGLAND: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 198810
AB Profile analysis measures the similarity between a target sequence and a group of **aligned** sequences (the probe). The probe sequences are used to produce a position-specific scoring table (the profile) that can be **aligned** with any sequence (the target) using standard dynamic programming methods. We are developing a library of profiles, each describing a different structural motif. This allows any target sequence to be rapidly scanned for the presence of structural motifs. Levels of significance for the comparison of target sequences with the profile are determined in advance, permitting an objective decision to be made as to whether a protein is likely to possess a structural motif.
CT Check Tags: Animal; Human; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
***Amino Acid Sequence**
Calcium-Binding Proteins: GE, genetics
Carrier Proteins: GE, genetics
Globin: GE, genetics
Immunoglobulins: GE, genetics
Molecular Sequence Data
***Pattern Recognition**
***Protein Conformation**
Sequence Homology, Nucleic Acid
***Software**
RN 9004-22-2 (Globin)
CN 0 (zinc-binding protein); 0 (Calcium-Binding Proteins); 0 (Carrier Proteins)

L45 ANSWER 30 OF 30 MEDLINE
AN 87260806 MEDLINE
DN 87260806
TI Profile analysis: detection of distantly related proteins.
AU Gribskov M; McLachlan A D; **Eisenberg D**
NC GM 31299 (NIGMS)
SO PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1987 Jul) 84 (13) 4355-8.
Journal code: PV3. ISSN: 0027-8424.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals; Cancer Journals
EM 198710
AB Profile analysis is a method for detecting distantly related proteins by sequence comparison. The basis for comparison is not only the customary Dayhoff mutational-distance matrix but also the results of structural studies and information implicit in the **alignments** of the sequences of families of similar proteins. This information is expressed in a position-specific scoring table (profile), which is created from a group of sequences previously **aligned** by structural or sequence similarity. The similarity of any other sequence (target) to the group of **aligned** sequences (probe) can be tested by comparing the target to

the profile using dynamic programming algorithms. The profile method differs in two major respects from methods of sequence comparison in common use: (i) Any number of known sequences can be used to construct the profile, allowing more information to be used in the testing of the target than is possible with pairwise **alignment** methods. (ii) The profile includes the penalties for insertion or deletion at each position, which allow one to include the probe secondary structure in the testing scheme. Tests with globin and immunoglobulin sequences show that profile analysis can distinguish all members of these families from all other sequences in a database containing 3500 protein sequences.
CT Check Tags: Comparative Study; Support, Non-U.S. Gov't; Support, U.S. Gov't, Non-P.H.S.; Support, U.S. Gov't, P.H.S.
Amino Acid Sequence
***Base Sequence**
Globin: GE, genetics
Immunoglobulins: GE, genetics
Information Systems
Protein Conformation
***Proteins: GE, genetics**
***Sequence Homology, Nucleic Acid**
Software
RN 9004-22-2 (Globin)

=> fil biosis

FILE 'BIOSIS' ENTERED AT 13:19:50 ON 12 FEB 2001
COPYRIGHT (C) 2001 BIOSIS(R)

FILE COVERS 1969 TO DATE.
CAS REGISTRY NUMBERS AND CHEMICAL NAMES (CNs) PRESENT FROM JANUARY 1969 TO DATE.

RECORDS LAST ADDED: 7 February 2001 (20010207/ED)

The BIOSIS file has been reloaded. Enter HELP RLOAD and HELP REINDEXING for details.

=> d all tot

L78 ANSWER 1 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 2001:62301 BIOSIS
DN PREV200100062301
TI From genome sequences to protein functions.
AU **Eisenberg, David (1); Marcotte, Edward (1); Pellegrini, Matteo (1); Thompson, Michael (1)**; Rotstein, Sergio (1); **Yeates, Todd (1)**
CS (1) UCLA, Los Angeles, CA, 90095-1570 USA
SO Biochemical Society Transactions, (October, 2000) Vol. 28, No. 5, pp. A120. print.
Meeting Info.: 18th International Congress of Biochemistry and Molecular Biology Birmingham, UK July 16-20, 2000
ISSN: 0300-5127.
DT **Conference**
LA English
SL English
CC Biochemical Studies - General *10060
General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals *00520
Mathematical Biology and Statistical Methods *04500
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Physiology and Biochemistry of Bacteria *31000
Plant Physiology, Biochemistry and Biophysics - Chemical Constituents *51522
BC Mycobacteriaceae 08881

IT Major Concepts
Biochemistry and Molecular Biophysics; Mathematical Biology (Computational Biology)
IT Chemicals & Biochemicals
proteins
IT Methods & Equipment
Rosetta Stone method: mathematical method; phylogenetic profile method: mathematical method
IT Miscellaneous Descriptors
genome sequences; Meeting Abstract
ORGN Super Taxa
Ascomycetes: Fungi, Plantae; Mycobacteriaceae: Mycobacteria, Actinomycetes and Related Organisms, Eubacteria, Bacteria, Microorganisms
ORGN Organism Name
Mycobacterium tuberculosis (Mycobacteriaceae); Saccharomyces cerevisiae (Ascomycetes)
ORGN Organism Superterms
Bacteria; Eubacteria; Fungi; Microorganisms; Nonvascular Plants; Plants

L78 ANSWER 2 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1998:56042 BIOSIS
DN PREV199800056042
TI Oligomer formation by 3D domain swapping: A model for protein **assembly** and misassembly.
AU Schlunegger, Michael P. (1); Bennett, Melanie J.; **Eisenberg, David (1)**
CS (1) Mol. Biol. Inst., Dep. Chem. and Biochem., Univ. Calif.-Los Angeles, Los Angeles, CA 90095-1570 USA
SO Wetzel, R. [Editor]. Advances in Protein Chemistry, (1997) Vol. 50, pp. 61-122. Advances in Protein Chemistry; Protein misassembly. Publisher: Academic Press, Inc. 1250 Sixth Ave., San Diego, California 92101, USA.
ISSN: 0065-3233. ISBN: 0-12-034250-2.
DT Book; General Review
LA English
CC **Biochemical Studies - Proteins, Peptides and Amino Acids *10064**
Genetics and Cytogenetics - General *03502
Comparative Biochemistry, General *10010
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biophysics - Molecular Properties and Macromolecules *10506
Metabolism - Proteins, Peptides and Amino Acids *13012
IT Major Concepts
Biochemistry and Molecular Biophysics: primary, primary
IT Chemicals & Biochemicals
amino acids; proteins: amino acid sequence, amino acid sequence
IT Miscellaneous Descriptors
mutations; protein **assembly** models; protein denaturation; protein misassembly models; three-dimensional domain swapping; Book Chapter

L78 ANSWER 3 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1997:448588 BIOSIS
DN PREV199799747791
TI **Predicting** protein secondary structure with probabilistic schemata of **evolutionarily** derived information.
AU **Thompson, Michael J.**; Goldstein, Richard A. (1)
CS (1) Dep. Chemistry, Univ. Michigan, Ann Arbor, MI 48109-1055 USA
SO Protein Science, (1997) Vol. 6, No. 9, pp. 1963-1975.
ISSN: 0961-8368.
DT Article
LA English
AB We demonstrate the applicability of our previously developed Bayesian probabilistic approach for predicting residue solvent accessibility to the problem of predicting secondary structure. Using only single-sequence data, this method achieves a three-state accuracy of 67% over a database

of 473 non-homologous proteins. This approach is more amenable to inspection and less likely to overlearn specifics of a dataset than "black box" methods such as neural networks. It is also conceptually simpler and less computationally costly. We also introduce a novel method for representing and incorporating multiple-sequence alignment information within the prediction algorithm, achieving 72% accuracy over a dataset of 304 non-homologous proteins. This is accomplished by creating a statistical model of the evolutionarily derived correlations between patterns of amino acid substitution and local protein structure. This model consists of parameter vectors, termed "substitution schemata," which probabilistically encode the structure-based heterogeneity in the distributions of amino acid substitutions found in alignments of homologous proteins. The model is optimized for structure prediction by maximizing the mutual information between the set of schemata and the database of secondary structures. Unlike "expert heuristic" methods, this approach has been demonstrated to work well over large datasets. Unlike the opaque neural network algorithms, this approach is physicochemically intelligible. Moreover, the model optimization procedure, the formalism for predicting one-dimensional structural features, and our previously developed method for tertiary structure recognition all share a common Bayesian probabilistic basis. This consistency starkly contrasts with the hybrid and ad hoc nature of methods that have dominated this field in recent years.

CC **Evolution *01500**
Mathematical Biology and Statistical Methods *04500
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Biocybernetics *10515

IT Major Concepts
Biochemistry and Molecular Biophysics; Evolution and Adaptation; Mathematical Biology (Computational Biology); Models and Simulations (Computational Biology)

IT Miscellaneous Descriptors
BAYESIAN STATISTICS; BIOCHEMISTRY AND BIOPHYSICS; MATHEMATICAL BIOLOGY; MOLECULAR EVOLUTION INFORMATION; PROBABILISTIC SCHEMATA; PROTEIN; PROTEIN SECONDARY STRUCTURE PREDICTION

L78 ANSWER 4 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1997:420310 BIOSIS
DN PREV199799719513
TI Predicting protein folds with the H3P2 method.
AU Rice, D. W.; **Eisenberg, D.**
CS UCLA, Los Angeles, CA USA
SO FASEB Journal, (1997) Vol. 11, No. 9, pp. A1036.
Meeting Info.: 17th International Congress of Biochemistry and Molecular Biology in conjunction with the Annual Meeting of the American Society for Biochemistry and Molecular Biology San Francisco, California, USA August 24-29, 1997
ISSN: 0892-6638.
DT **Conference**; Abstract
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Mathematical Biology and Statistical Methods *04500
Comparative Biochemistry, General *10010
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506

IT Major Concepts
Biochemistry and Molecular Biophysics; Mathematical Biology (Computational Biology); Methods and Techniques

IT Miscellaneous Descriptors
AMINO ACIDS; BIOCHEMISTRY AND BIOPHYSICS; DYNAMIC PROGRAMMING ALGORITHM; H3P2 MATRIX; H3P2 METHOD; METHODOLOGY; MISCELLANEOUS METHOD; PROTEIN FOLDING PREDICTION; PROTEINS

L78 ANSWER 5 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1997:251190 BIOSIS
DN PREV199799550493
TI A 3D-1D substitution matrix for protein fold **recognition** that includes **predicted** secondary structure of the sequence.
AU Rice, Danny W.; **Eisenberg, David (1)**
CS (1) UCLA-DOE Lab. Structural Biol. Mol. Med., Mol. Biol. Inst., UCLA, Box 951570, Los Angeles, CA 90095-1570 USA
SO Journal of Molecular Biology, (1997) Vol. 267, No. 4, pp. 1026-1038. ISSN: 0022-2836.
DT Article
LA English
AB In protein fold recognition, a probe amino acid sequence is compared to a library of representative folds of known structure to identify a structural homolog. In cases where the probe and its homolog have clear sequence similarity, traditional residue substitution matrices have been used to predict the structural similarity. In cases where the probe is sequentially distant from its homolog, we have developed a (7 times 3 times 2 times 7 times 3) 3D-1D substitution matrix (called H3P2), calculated from a database of 116 structural pairs. Members of each pair share a similar fold, but have sequence identity less than 30%. Each probe sequence position is defined by one of seven residue classes and three secondary structure classes. Each homologous fold position is defined by one of seven residue classes, three secondary structure classes, and two burial classes. Thus the matrix is five-dimensional and contains 7 times 3 times 2 times 7 times 3 = 882 elements or 3D-1D scores. The first step in assigning a probe sequence to its homologous fold is the prediction of the three-state (helix, strand, coil) secondary structure of the probe; here we use the profile based neural network prediction of secondary structure (PHD) program. Then a dynamic programming algorithm uses the H3P2 matrix to align the probe sequence with structures in a representative fold library. To test the effectiveness of the H3P2 matrix a challenging, fold class diverse, and cross-validated benchmark assessment is used to compare the H3P2 matrix to the GONNET, PAM250, BLOSUM62 and a secondary structure only substitution matrix. For distantly related sequences the H3P2 matrix detects more homologous structures at higher reliabilities than do these other substitution matrices, based on sensitivity versus specificity plots (or SENS-SPEC plots). The added efficacy of the H3P2 matrix arises from its information on the statistical preferences for various sequence-structure environment combinations from very distantly related proteins. It introduces the predicted secondary structure information from a sequence into fold recognition in a statistical way that normalizes the inherent correlations between residue type, secondary structure and solvent accessibility.
CC General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Mathematical Biology and Statistical Methods *04500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Information Studies; Mathematical Biology (Computational Biology); Methods and Techniques
IT Miscellaneous Descriptors
BIOCHEMISTRY AND BIOPHYSICS; COMPUTER METHOD; COMPUTER PROGRAM; DYNAMIC PROGRAMMING ALGORITHM; FOLD RECOGNITION; MATHEMATICAL BIOLOGY; MATHEMATICAL METHOD; PROFILE BASED NEURAL NETWORK PREDICTION OF SECONDARY STRUCTURE; PROTEIN; SECONDARY STRUCTURE PREDICTION; SOLVENT ACCESSIBILITY; THREE-DIMENSIONAL-ONE-DIMENSIONAL STRUCTURE SUBSTITUTION MATRIX

L78 ANSWER 6 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1997:126718 BIOSIS
DN PREV199799418531
TI Into the black of night. (Second Meeting on Critical Assessment of Techniques For Protein Structure Prediction, Asilomar, California, USA,

December 12-17, 1996.
AU **Eisenberg, David**
CS UCLA-DOE Lab., Structural Biol. Mol. Med., Box 95157, UCLA, Los Angeles, CA 90095-1570 USA
SO Nature Structural Biology, (1997, Vol. 4, No. 2, pp. 95-97.
ISSN: 1072-8368.
DT **Conference**; Report
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Methods and Techniques
IT Miscellaneous Descriptors
ASILOMAR; BIOCHEMISTRY AND BIOPHYSICS; BOLTZMANN; COMPUTATIONAL BIOCHEMISTRY; DARWIN; FOLDING; PREDICTION METHODS; PROTEIN; PROTEIN FOLDING; PROTEIN STRUCTURE; PROTEIN STRUCTURE PREDICTION TECHNIQUES; SCHRODINGER; STRUCTURE
GT California (USA, North America, Nearctic region); USA (North America, Nearctic region)

L78 ANSWER 7 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1996:311111 BIOSIS
DN PREV199699033806
TI **Predicting** solvent accessibility: Higher accuracy using Bayesian statistics and optimized residue substitution classes.
AU **Thompson, Michael J.**; Goldstein, Richard A. (1)
CS (1) Biophysics Res. Div. Dep. Chemistry, Univ. Michigan, Ann Arbor, MI 48109-1055 USA
SO Proteins Structure Function and Genetics, (1996) Vol. 25, No. 1, pp. 38-47.
ISSN: 0887-3585.
DT Article
LA English
AB We introduce a novel Bayesian probabilistic method for predicting the solvent accessibilities of amino acid residues in globular proteins. Using single sequence data, this method achieves prediction accuracies higher than previously published methods. Substantially improved predictions-comparable to the highest accuracies reported in the literature to date-are obtained by representing alignments of the example proteins and their homologs as strings of residue substitution classes, depending on the side chain types observed at each alignment position. These results demonstrate the applicability of this relatively simple Bayesian approach to structure prediction and illustrate the utility of the classification methodology previously developed to extract information from aligned sets of structurally related proteins.
CC **Mathematical Biology and Statistical Methods *04500**
Biochemical Studies - General 10060
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Mathematical Biology (Computational Biology)
IT Miscellaneous Descriptors
AMINO ACID SUBSTITUTION; GLOBULAR PROTEIN

L78 ANSWER 8 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1996:278170 BIOSIS
DN PREV199699030626
TI Protein fold **recognition** using sequence-derived **predictions**.
AU Fischer, Daniel; **Eisenberg, David (1)**
CS (1) UCLA-DOE Lab. Structural Biol. Molecular Med., Molecular Biol. Inst., UCLA, Box 951570, Los Angeles, CA 90095-1570 USA

SO Protein Science, (1996) Vol. 5, No. 5, pp. 947-955.
ISSN: 0961-8368.
DT Article
LA English
AB In protein fold recognition, one assigns a probe amino acid sequence of unknown structure to one of a library of target 3D structures. Correct assignment depends on effective scoring of the probe sequence for its compatibility with each of the target structures. Here we show that, in addition to the amino acid sequence of the probe, sequence-derived properties of the probe sequence (such as the predicted secondary structure) are useful in fold assignment. The additional measure of compatibility between probe and target is the level of agreement between the predicted secondary structure of the probe and the known secondary structure of the target fold. That is, we recommend a sequence-structure compatibility function that combines previously developed compatibility functions (such as the 3D-1D scores of Bowie et al. (1991) or sequence-sequence replacement tables) with the predicted secondary structure of the probe sequence. The effect on fold assignment of adding predicted secondary structure is evaluated here by using a benchmark set of proteins (Fischer et al., 1996a). The 3D structures of the probe sequences of the benchmark are actually known, but are ignored by our method. The results show that the inclusion of the predicted secondary structure improves fold assignment by about 25%. The results also show that, if the true secondary structure of the probe were known, correct fold assignment would increase by an additional 5-32%. We conclude that incorporating sequence-derived predictions significantly improves assignment of sequences to known 3D folds. Finally, we apply the new method to assign folds to sequences in the SWISSPROT database; six fold assignments are given that are not detectable by standard sequence-sequence comparison methods; for two of these, the fold is known from X-ray crystallography and the fold assignment is correct.
CC **Mathematical Biology and Statistical Methods *04500**
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Mathematical Biology (Computational Biology); Methods and Techniques
IT Miscellaneous Descriptors
ANALYTICAL METHOD; BIOCHEMISTRY AND MOLECULAR BIOPHYSICS; PROTEIN FOLD RECOGNITION METHOD; PROTEIN PROBE; SECONDARY STRUCTURE; SEQUENCE-STRUCTURE COMPATIBILITY FUNCTION

L78 ANSWER 9 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1996:100655 BIOSIS
DN PREV199698672790
TI Three-dimensional profiles: Assessing the compatibility of an amino acid sequence with a three-dimensional structure.
AU Bowie, James U. (1); Luthy, Roland; **Eisenberg, David**
CS (1) Molecular Biol. Inst., Univ. Calif. at Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90024-1570 USA
SO Go, M. [Editor]; Schimmel, P. [Editor]. (1995) pp. 297-309. Tracing biological evolution in protein and gene structures.
Publisher: Elsevier Science Publishers B.V. PO Box 211, Sara Burgerhartstraat 25, 1000 AE Amsterdam, Netherlands.
Meeting Info.: 20th Taniguchi International Symposium, Division of Biophysics Nagoya, Japan October 31-November 4, 1994
ISBN: 0-444-82187-2.
DT Book; **Conference**
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Evolution *01500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506

BC *04500
IT Major Concepts
Biochemistry and Molecular Biophysics; Evolution and Adaptation; Methods and Techniques
IT Miscellaneous Descriptors
ANALYTICAL METHOD; BOOK CHAPTER; MEETING PAPER; MOLECULAR EVOLUTION; MOLECULAR MODELLING; PROTEIN FOLDING
ORGN Organism Name
organism (Organisms - Unspecified); organisms (Organisms - Unspecified)

L78 ANSWER 10 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1995:142564 BIOSIS
DN PREV199698857864
TI Inverse protein folding by the residue pair preference **profile** method: Estimating the correctness of alignments of structurally compatible sequences.
AU Wilmanns, Matthias (1); **Eisenberg, David**
CS (1) European Molecular Biol. Lab., Meyerhofstr. 1, Postfach 10.2209, D-69012 Heidelberg Germany
SO Protein Engineering, (1995) Vol. 8, No. 7, pp. 627-639.
ISSN: 0269-2139.
DT Article
LA English
AB The residue pair preference profile (R3P) method is an inverse folding method that combines environmental profiles and pair preference profiles. The method uses statistical preferences for residue pairs which score the likelihood of finding a profiled residue to be paired with a residue within its local environment. All pairs are characterized by their dihedral angles, secondary structure and number of neighboring residues as a function of residue type. Each residue pair preference is expressed for all 20 amino acids of the profiled residue and is weighted by the compatibility of the environment residue with its own local environment. The R3P method produces an initial profile sequence alignment which is then refined by converting the initial profile into a profile of a target sequence threaded into the structure of the initial profile. We have tested this method by evaluating alignments of sequences with known 3-D structures using structural superposition alignments as reference. R3P-sequence alignments are gtoreq 50% correct on average for sequences whose 3-D structure pairs superimpose with an r.m.s. deviation of ltoreq 1.97 ANG . The average improvement in correctness during this iterative refinement is 14%. The R3P-sequence alignments are compared with sequence-sequence and 3-D profile-sequence alignments. When all three methods are combined, on average gtoreq 50% of the alignments are correct for pairs of 3-D structures that superimpose within 2.12 ANG . A 3-D model of HisA is predicted with the combined method.
CC **Mathematical Biology and Statistical Methods 04500**
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Biocybernetics *10513
IT Major Concepts
Biochemistry and Molecular Biophysics; Methods and Techniques; Models and Simulations (Computational Biology)
IT Miscellaneous Descriptors
ALPHA-HELIX; ANALYTICAL METHOD; BETA STRAND; MATHEMATICAL MODEL

L78 ANSWER 11 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1995:479129 BIOSIS
DN PREV199598493436
TI Local moves: An efficient **algorithm** for simultaneous of protein folding.
AU Elofsson, Arne; Le Grand, Scot M.; **Eisenberg, David (1)**
CS (1) UCLA-DOE Lab Structural Biol. Mol. Med., Mol. Biol. Inst., UCLA, Los Angeles, CA 90095-1570 USA
SO Proteins Structure Function and Genetics, (1995) Vol. 23, No. 1, pp. 73-82.

ISSN: 0887-3585.
DT Article
LA English
AB We have enhanced genetic algorithms and Monte Carlo methods for simulation of protein folding by introducing "local moves" in dihedral space. A local move consists of changes in backbone dihedral angles in a sequential window while the positions of all atoms outside the window remain unchanged. We find three advantages of local moves: (1) For some energy functions, protein conformations of lower energy are found; (2) these low energy conformations are found in fewer steps; and (3) the simulations are less sensitive to the details of the annealing protocol. To distinguish the effectiveness of local move algorithm from the complexity of the energy function, we have used several different energy functions. These energy functions include the Profile score (Bowie et al., Science 253:164-170, 1991), the knowledge-based energy function used by Bowie and Eisenberg 1994 (Proc. Natl. Acad. Sci. U.S.A. 91:4436-4440, 1994), two energy terms developed as suggested by Sippl and coworkers (Hendlich et al., J. Mol. Biol. 216:167-180, 1990), and AMBER (Weiner and Kollman, J. Comp. Chem. 2:287-303, 1981). Besides these energy functions we have used three energy functions that include knowledge of the native structures: the RMSD from the native structure, the distance matrix error, and an energy term based on the distance between different residue types called DBIN. In some of these simulations the main advantage of local moves is the reduced dependence on the details of the annealing schedule. In other simulations, local moves are superior to other algorithms as structures with lower energy are found.
CC **Mathematical Biology and Statistical Methods *04500**
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Mathematical Biology (Computational Biology); Methods and Techniques
IT Miscellaneous Descriptors
DIHEDRAL ANGLES; MATHEMATICAL METHOD; MONTE CARLO SIMULATION; PROTEIN STRUCTURE; RING CLOSURE; STRUCTURE PREDICTION

L78 ANSWER 12 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1995:138147 BIOSIS
DN PREV199598152447
TI An iconography of amino acids based on mutual information between protein sequence and local structure.
AU **Thompson, M. J. (1)**; Goldstein, R. A.
CS (1) Biophysics Res. Div., Univ. Mich., Ann Arbor, MI 48109-1055 USA
SO Biophysical Journal, (1995) Vol. 68, No. 2 PART 2, pp. A115.
Meeting Info.: 39th Annual Meeting of the Biophysical Society San Francisco, California, USA February 12-16, 1995
ISSN: 0006-3495.
DT **Conference**
LA English
CC **Evolution *01500**
Mathematical Biology and Statistical Methods *04500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Evolution and Adaptation; Mathematical Biology (Computational Biology); Methods and Techniques
IT Miscellaneous Descriptors
DESCRIPTION CLASS; EVOLUTION; MATHEMATICAL METHOD; MEETING ABSTRACT; METROPOLIS SEARCH SCHEME; SECONDARY STRUCTURE; STATISTICS; STRUCTURAL PATTERN; TERTIARY STRUCTURE

L78 ANSWER 13 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1994:341834 BIOSIS
DN PREV199497354834

TI Verification of protein structures: **Patterns** of nonbonded atomic interactions.
AU Colovos, Chris; **Yeates, Todd O. (1)**
CS (1) Dep. Chem. and Biochem., Univ. Calif., 405 Hilgard Avenue, Los Angeles, CA 90024-1569 USA
SO Protein Science, (1993) Vol. 2, No. 9, pp. 1511-1519.
ISSN: 0961-8368.
DT Article
LA English
AB A novel method for differentiating between correctly and incorrectly determined regions of protein structures based on characteristic atomic interactions is described. Different types of atoms are distributed nonrandomly with respect to each other in proteins. Errors in model building lead to more randomized distributions of the different atom types, which can be distinguished from correct distributions by statistical methods. Atoms are classified in one of three categories: carbon (C), nitrogen (N), and oxygen (O). This leads to six different combinations of pairwise noncovalently bonded interactions (CC, CN, CO, NN, NO, and OO). A quadratic error function is used to characterize the set of pairwise interactions from nine-residue sliding windows in a database of 96 reliable protein structures. Regions of candidate protein structures that are mistraced or misregistered can then be identified by analysis of the pattern of nonbonded interactions from each window.
CC General Biology - Information, Documentation, Retrieval and Computer Applications *00510
Mathematical Biology and Statistical Methods *04500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Biocybernetics *10510
IT Major Concepts
Biochemistry and Molecular Biophysics; Information Studies; Mathematical Biology (Computational Biology); Methods and Techniques; Models and Simulations (Computational Biology)
IT Miscellaneous Descriptors
ANALYTICAL METHOD; CRYSTAL STRUCTURE; IMAGES ON DISKETTE; KINEMAGE; MATHEMATICAL METHOD; MATHEMATICAL MODEL

L78 ANSWER 14 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1994:295332 BIOSIS
DN PREV199497308332
TI An **evolutionary** approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function.
AU Bowie, James U. (1); **Eisenberg, David**
CS (1) Dep. Chem. Biochem., Univ. California Los Angeles-Dep. Energy Lab., Structural Biol., Univ. California, Los Angeles, CA 90024-1570 USA
SO Proceedings of the National Academy of Sciences of the United States of America, (1994) Vol. 91, No. 10, pp. 4436-4440.
ISSN: 0027-8424.
DT Article
LA English
AB Three short protein sequences have been guided by computer to folds resembling their crystal structures. Initially, peptide fragment conformations ranging in size from 9 to 25 residues were selected from a database of known protein structures. A fragment was selected if it was compatible with a segment of the sequence to be folded, as judged by three-dimensional profile scores. By linking the selected fragment conformations together, hundreds of trial structures were generated of the same length and sequence as the protein to be folded. These starting trial structures were then improved by an evolutionary algorithm. Selection pressure for improving the structures was provided by an energy function that was designed to guide the conformational search procedure toward the correct structure. We find that by evolution of only 400 structures for fewer than 1400 generations, the overall fold of some small helical proteins can be computed from the sequence, with deviations from observed structures of 2.5-4.0 ANG for C-alpha atoms.

CC General Biology - Information, Documentation, Retrieval and Computer Applications 00530
Mathematical Biology and Statistical Methods 04500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids 10064
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Biocybernetics *10515
IT Major Concepts
Biochemistry and Molecular Biophysics; Methods and Techniques
IT Miscellaneous Descriptors
ALGORITHM; COMPUTER; CONFORMATION; CRYSTAL STRUCTURE; PROTEIN FOLDING

L78 ANSWER 15 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1994:287246 BIOSIS
DN PREV199445005366
TI 3D profiles: Principles and applications.
AU Bowie, James U.; Luthy, Roland; Wilmanns, Matthias; Wesson, Laura; Zhang, Kam; **Eisenberg, David**
CS Mol. Biol. Inst., Los Angeles, CA 90024-1570 USA
SO Protein Engineering, (1993) Vol. 6, No. SUPPL., pp. 123.
Meeting Info.: Winter Symposium on Advances in Gene Technology: Protein Engineering and Beyond Miami, Florida, USA 1993
ISSN: 0269-2139.
DT **Conference**
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Mathematical Biology and Statistical Methods *04500
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Major Concepts
Biochemistry and Molecular Biophysics; Mathematical Biology (Computational Biology)
IT Miscellaneous Descriptors
ABSTRACT; MOLECULAR DYNAMICS; PROTEIN FOLDING; TERTIARY CONFORMATION; 3-DIMENSIONAL

L78 ANSWER 16 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1993:270279 BIOSIS
DN PREV199396000504
TI **Computer simulation** of antibody binding specificity.
AU **Pellegrini, Matteo**; Doniach, Sebastian (1)
CS (1) Dep. Applied Physics, Stanford Univ., Stanford, CA 94305-4090 USA
SO Proteins Structure Function and Genetics, (1993) Vol. 15, No. 4, pp. 436-444.
ISSN: 0887-3585.
DT Article
LA English
AB A Monte Carlo algorithm that searches for the optimal docking configuration of hen egg white lysozyme to an antibody is developed. Both the lysozyme and the antibody are kept rigid. Unlike the work of other authors, our algorithm does not attempt to explicitly maximize surface contact, but minimizes the energy computed using coarse-grained pair potentials. The final refinement of our best solutions using all-atom OPLS potentials (Jorgensen and Tirado-Rives-8) consistently yields the native conformation as the preferred solution for three different antibodies. We find that the use of an exponential distance-dependent dielectric function is an improvement over the more commonly used linear form.
CC General Biology - Information, Documentation, Retrieval and Computer Applications 00530
Mathematical Biology and Statistical Methods 04500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Methods - Carbohydrates *10058
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biochemical Studies - Carbohydrates *10068
Biophysics - Biocybernetics *10515

Enzymes - Physiological Studies *10809
BC Galliformes *85536
IT Major Concepts
Biochemistry and Molecular Biophysics; Methods and Techniques; Models and Simulations (Computational Biology)
IT Chemicals & Biochemicals
LYSOZYME
IT Miscellaneous Descriptors
LYSOZYME; MATHEMATICAL MODEL
ORGN Super Taxa
Galliformes: Aves, Vertebrata, Chordata, Animalia
ORGN Organism Name
chicken (Galliformes)
ORGN Organism Superterms
animals; birds; chordates; nonhuman vertebrates; vertebrates
RN 9001-63-2 (LYSOZYME)

L78 ANSWER 17 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1993:155005 BIOSIS
DN PREV199344073805
TI Three-dimensional profiles for analysing protein sequence-structure relationships.
AU **Eisenberg, David**; Bowie, James U.; Luthy, Roland; Choe, Seunghyun
CS Molecular Biol. Inst., Dep. Chem. Biochemistry, UCLA, Los Angeles, CA 90024
SO Sarre, P. J. (Editor). Faraday Discussions, (1992) No. 93, pp. 25-34. Faraday Discussions; Structure and activity of enzymes. Publisher: Royal Society of Chemistry Piccadilly House, London W1V 0BN, England.
Meeting Info.: Meeting Cambridge, England, UK April 1-3, 1991
ISSN: 1359-6640. ISBN: 0-85186-425-..
DT Article
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Enzymes - Chemical and Physical *10806
Physiology and Biochemistry of Bacteria *31000
BC Irregular Nonsporing Gram-Positive Rods *06890
IT Major Concepts
Biochemistry and Molecular Biophysics; Enzymology (Biochemistry and Molecular Biophysics); Physiology
IT Chemicals & Biochemicals
DIPHTHERIA TOXIN
IT Sequence Data
protein sequence
IT Miscellaneous Descriptors
ANALYSING; DIPHTHERIA TOXIN; OLIGOMERIC STATE; PROTEIN MODEL; X-RAY STRUCTURE
ORGN Super Taxa
Irregular Nonsporing Gram-Positive Rods: Eubacteria, Bacteria
ORGN Organism Name
irregular nonsporing gram-positive rods (Irregular Nonsporing Gram-Positive Rods)
ORGN Organism Superterms
bacteria; eubacteria; microorganisms
RN 58917-16-1 (DIPHTHERIA TOXIN)

L78 ANSWER 18 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1991:369586 BIOSIS
DN BA92:48811
TI SECONDARY STRUCTURE-BASED **PROFILES** USE OF STRUCTURE-CONSERVING SCORING TABLES IN SEARCHING PROTEIN SEQUENCE DATABASES FOR STRUCTURAL SIMILARITIES.

AU LUTHY R; MCLACHLAN A D; **EISENBERG D**
CS MOL. BIOL. INST., UNIV. CALIF. AT LOS ANGELES, LOS ANGELES, CALIF. 90024-1570, USA.
SO PROTEINS STRUCT FUNCT GENET, (1991) 10 (3), 229-239.
CODEN: PSFGEY. ISSN: 0887-3585.
FS BA; OLD
LA English
AB The profile method, for detecting distantly related proteins by sequence comparison, has been extended to incorporate secondary structure information from known X-ray structures. The sequence of a known structure is aligned to sequences of other members of a given folding class. From the known structure, the secondary structure (.alpha.-helix, .beta.-strand or "other") is assigned to each position of the aligned sequences. As in the standard profile method, a position-dependent scoring table, termed a profile, is calculated from the aligned sequences. However, rather than using the standard Dayhoff mutation table in calculating the profile, we use distinct amino acid mutation tables for residues in .alpha.-helices, .beta.-late the profile. In addition, we also distinguish between internal and external residues. With this new secondary structure-based profile method, we created a profile for eight-stranded, antiparallel .beta. barrels of the insecticyanin folding class. It is based on the sequences of retinol-binding protein, insecticyanin and .beta.-lactoglobulin. Scanning the sequence database with this profile, it was possible to detect the sequence of avidin. The structure of streptavidin is known, and it appears to be distantly related to the antiparallel .beta. barrels. Also detected is the sequence of complement component C8, which we therefore predict to be a member of this folding class.
CC General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Genetics and Cytogenetics - General *03501
Mathematical Biology and Statistical Methods *04500
Radiation - Radiation and Isotope Techniques 06504
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
AMINO ACID MUTATION TABLE X-RAY STRUCTURES

L78 ANSWER 19 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1991:355290 BIOSIS
DN BR41:39905
TI 3-D PROFILES.
AU **EISENBERG D**; BOWIE J; LUTHY R
CS MOL. BIOL. INST., UNIV. CALIF., LOS ANGELES, CALIF. 90024-1570.
SO MEETING ON PROTEIN FOLDING, STRUCTURE AND FUNCTION HELD AT THE 20TH ANNUAL MEETING OF THE KEYSTONE SYMPOSIA ON MOLECULAR AND CELLULAR BIOLOGY, KEYSTONE, COLORADO, USA, APRIL 8-14, 1991. J CELL BIOCHEM SUPPL. (1991) 0 (15 PART G), 166.
CODEN: JCBSD7.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Mathematical Biology and Statistical Methods *04500
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT PROTEIN SEQUENCE ANALYSIS GLOBINS PROTEIN FOLDING COMPUTER PROGRAM

L78 ANSWER 20 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1991:59406 BIOSIS
DN BR40:24821

TI **PROFILE** ANALYSIS.
AU GRIBSKOV M; LUTHY R; **EISENBERG D**
CS BRI-BASIC RES. PROGRAM, NATL. CANCER INST.-FREDERICK CANCER RES. FACILITY, FREDERICK, MD. 21701.
SO DOOLITTLE, R. F. (ED.). METHODS IN ENZYMOLOGY, VOL. 183. MOLECULAR EVOLUTION: COMPUTER ANALYSIS OF PROTEIN AND NUCLEIC ACID SEQUENCES. XXIX+736P. ACADEMIC PRESS, INC.: SAN DIEGO, CALIFORNIA, USA; LONDON, ENGLAND, UK. ILLUS. (1990) 0 (0), 146-166.
CODEN: MENZAU. ISSN: 0076-6879. ISBN: 0-12-182084-X.
FS RF; OLD
LA English
CC General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Mathematical Biology and Statistical Methods *04500
Comparative Biochemistry, General *10010
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
IT Miscellaneous Descriptors
PROTEIN SEQUENCE DATABASES

L78 ANSWER 21 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1990:254277 BIOSIS
DN BR38:119065
TI DETECTION OF PROTEIN STRUCTURAL FEATURES WITH PROFILE ANALYSIS.
AU GRIBSKOV M; **EISENBERG D**
CS NCI-FREDERICK CANCER RES. FAC., BRI-BASIC RES. PROGRAM, FREDERICK, MD. 21701-1013, USA.
SO HUGLI, T. E. (ED.). TECHNIQUES IN PROTEIN CHEMISTRY; SECOND ANNUAL MEETING, SAN DIEGO, CALIFORNIA, USA, AUGUST 13-17, 1988. XVII+612P. ACADEMIC PRESS, INC.: SAN DIEGO, CALIFORNIA, USA; LONDON, ENGLAND, UK. ILLUS. (1989) 0 (0), 108-118.
ISBN: 0-12-68200I-5 (PAPER), 0-12-68200-7 (CLOTH).
DT **Conference**
FS RF; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
SEQUENCE HOMOLOGIES 3-DIMENSIONAL STRUCTURES COMPUTER PROGRAMS

L78 ANSWER 22 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1990:208405 BIOSIS
DN BR38:97023
TI STRUCTURE OF ALPHA 1-12 A DESIGNED SYNTHETIC PROTEIN MODEL.
AU **EISENBERG D**; HILL C P; ANDERSON D H; WESSON M
CS MOL. BIOL. INST., UCLA, LOS ANGELES, CALIF. 90024.
SO THIRTY-FOURTH ANNUAL MEETING OF THE BIOPHYSICAL SOCIETY, BALTIMORE, MARYLAND, USA, FEBRUARY 18-22, 1990. BIOPHYS J. (1990) 57 (2 PART 2), 384A.
CODEN: BIOJAU. ISSN: 0006-3495.
DT **Conference**
FS RF; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - General Biophysical Techniques 10504
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT

L78 ANSWER 23 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1990:10[illegible]37 BIOSIS
DN BF38:48[illegible]22
TI INTERPRETATION OF PROTEIN FOLDING AND BINDING WITH ATOMIC SOLVATION PARAMETERS.
AU **EISENBERG D**; WESSON M; YAMASHITA M
CS MOL. BIOL. INST., UNIV. CALIF., LOS ANGELES, CALIF. 90024-1570, USA.
SO NOBEL SYMPOSIUM ON "STRUCTURE AND DYNAMICS IN BIOLOGICAL SYSTEMS," LUND, (SWEDEN), SWEDEN, DECEMBER 6-9, 1988. CHEM SCR. (1989) 29A (C), 217-222.
CODEN: CSPPB9. ISSN: 0004-2056.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals** 00520
Biochemical Methods - Proteins, Peptides and Amino Acids 10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - General Biophysical Techniques *10504
Biophysics - Molecular Properties and Macromolecules *10506

L78 ANSWER 24 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1988:477624 BIOSIS
DN BF35:107514
TI APPLICATIONS OF HYDROPHOBIC MOMENTS AND ATOMIC SOLVATION PARAMETERS ASPS TO UNDERSTANDING PROTEIN FOLDING AND LIGAND BINDING.
AU **EISENBERG D**; WESSON M; YAMASHITA M
CS MOLECULAR BIOL. INST., LOS ANGELES, CA 90024.
SO 154TH NATIONAL AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE ANNUAL MEETING, BOSTON, MASSACHUSETTS, USA, FEBRUARY 11-15, 1988. AM ASSOC ADV SCI ABSTR PAP NATL MEET. (1988) 0 (154), 35.
CODEN: ANMSD6.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals** 00520
Comparative Biochemistry, General 10010
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - General Biophysical Studies *10502
Biophysics - General Biophysical Techniques 10504
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Membrane Phenomena *10508
IT Miscellaneous Descriptors
ABSTRACT MEMBRANE SEGMENTS HELIX

L78 ANSWER 25 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1988:261316 BIOSIS
DN BA86:560
TI **PROFILE** SCANNING FOR THREE-DIMENSIONAL STRUCTURAL **PATTERNS** IN PROTEIN SEQUENCES.
AU GRIBSKOV M; HOMYAK M; EDENFIELD J; **EISENBERG D**
CS MOL. BIOL. INST., 405 HILGARD AVE., UNIV. CALIF., LOS ANGELES, CALIF. 90024-1570, USA.
SO COMPUT APPL BIOSCI, (1988) 4 (1), 61-66.
CODEN: COABEF. ISSN: 0266-7061.
FS BA; OLD
LA English
AB Profile analysis measures the similarity between a target sequence and a group of aligned sequences (the probe). The probe sequences are used to produce a position-specific scoring table (the profile) that can be aligned with any sequence (the target) using standard dynamic programming methods. We are developing a library of profiles, each describing a different structural motif. This allows any target sequence to be rapidly scanned for the presence fo structural motifs. Levels of significance for

the comparison of target sequences with the profile are determined in advance, permitting an objective decision to be made as to whether a protein is likely to possess a structural motif.

CC General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Mathematical Biology and Statistical Methods 04500
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Methods - Carbohydrates 10058
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biochemical Studies - Carbohydrates 10068
Biophysics - Molecular Properties and Macromolecules *10506

IT Miscellaneous Descriptors
IMMUNOGLOBULIN ALGORITHM GLOBIN PROFILE

L78 ANSWER 26 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1988:233376 BIOSIS
DN BR34:115896
TI STRUCTURAL STUDIES OF DIPHTHERIA TOXIN.
AU KANTARDJIEFF K; DIJKSTRA B; WESTBROOK E M; BARBIERI J T; CARROLL S F; COLLIER R J; **EISENBERG D**
CS MOL. BIOL. INST., UNIV. CALIFORNIA, LOS ANGELES, CALIF. 90024.
SO OXENDER, D. L. (ED.). UCLA (UNIVERSITY OF CALIFORNIA-LOS ANGELES) **SYMPOSIA** ON MOLECULAR AND CELLULAR BIOLOGY NEW SERIES, VOL. 69. PROTEIN STRUCTURE, FOLDING, AND DESIGN 2; DUPONT-UCLA **SYMPOSIUM**, STEAMBOAT SPRINGS, COLORADO, USA, APRIL 4-11, 1987. XXVI+550P. ALAN R. LISS, INC.: NEW YORK, NEW YORK, USA. ILLUS. (1987) 0 (0), 187-200. CODEN: USMBD6. ISSN: 0735-9543. ISBN: 0-8451-2668-7.
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Methods - Proteins, Peptides and Amino Acids 10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Physiology and Biochemistry of Bacteria *31000
BC Coryneform Group of Bacteria 05814
RN 58517-16-1 (DIPHTHERIA TOXIN)

L78 ANSWER 27 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1987:397203 BIOSIS
DN BA84:73383
TI **PROFILE** ANALYSIS DETECTION OF **DISTANTLY RELATED** PROTEINS.
AU GRIBSKOV M; MCLACHLAN A D; **EISENBERG D**
CS MOLECULAR BIOL. INST., DEP. CHEM. BIOCHEM., UNIV. CALIF., LOS ANGELES, CA 90024.
SO PROC NATL ACAD SCI U S A, (1987) 84 (13), 4355-4358. CODEN: PNASA6. ISSN: 0027-8424.
FS BA; OLD
LA English
AB Profile analysis is a method for detecting distantly related proteins by sequence comparison. The basis for comparison is not only the customary Dayhoff mutational-distance matrix but also the results of structural studies and information implicit in the alignments of the sequences of families of similar proteins. This information is expressed in a position-specific scoring table (profile), which is created from a group of sequences previously aligned by structural or sequence similarity. The similarity of any other sequence (target) to the group of aligned sequences (probe) can be tested by comparing the target to the profile using dynamic programming algorithms. The profile method differs in two major respects from methods of sequence comparison in common use: (i) Any number of known sequences can be used to construct the profile, allowing more information to be used in the testing of the target than is possible with pairwise alignment methods. (ii) The profile includes the penalties for insertion or deletion at each position, which allow one to include the probe secondary structure in the testing scheme. Tests with globin and

immunoglobulin sequences show that profile analysis can distinguish all members of these families from all other sequences in a database containing 3800 protein sequences.

CC General Biology - Information, Documentation, Retrieval and Computer Applications *00539
Genetics and Cytogenetics - General 03502
Mathematical Biology and Statistical Methods 04500
Comparative Biochemistry, General *10010
Biochemical Methods - Proteins, Peptides and Amino Acids 10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Immunology and Immunochemistry - General; Methods *34502

IT Miscellaneous Descriptors
GLOBIN IMMUNOGLOBULIN DYNAMIC PROGRAMMING ALGORITHMS SEQUENCE COMPARISON

L78 ANSWER 28 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1987:370062 BIOSIS
DN BR33:60537
TI THE DESIGN SYNTHESIS AND CHARACTERIZATION OF A FOUR-HELICAL BUNDLE PROTEIN.
AU DEGRADO W F; HO S P; WEBER P C; WASSERMAN Z R; **EISENBERG D**
CS E. I. DU PONT DE NEMOURS AND CO., CENTRAL RES. DEV. DEP., WILMINGTON, DE 19898, USA.
SO SYMPOSIUM ON PROTEIN STRUCTURE AND DESIGN HELD AT THE 16TH ANNUAL MEETING OF THE UCLA (UNIVERSITY OF CALIFORNIA-LOS ANGELES) SYMPOSIA ON MOLECULAR AND CELLULAR BIOLOGY, LOS ANGELES, CALIFORNIA, USA, APRIL 4-11, 1987. J CELL BIOCHEM SUPPL. (1987) 0 (11 PART C), 196.
CODEN: JCBSD7.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Methods - Proteins, Peptides and Amino Acids 10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT

L78 ANSWER 29 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1987:370057 BIOSIS
DN BR33:60532
TI PROTEIN FOLDING THREE APPROACHES.
AU **EISENBERG D**; ALMASSY R J; JANSON C; GRIBSKOV M; YAMASHITA M; WESSON M; REES D C
CS UNIV. CALIF., LOS ANGELES, CA 90024, USA.
SO SYMPOSIUM ON PROTEIN STRUCTURE AND DESIGN HELD AT THE 16TH ANNUAL MEETING OF THE UCLA (UNIVERSITY OF CALIFORNIA-LOS ANGELES) SYMPOSIA ON MOLECULAR AND CELLULAR BIOLOGY, LOS ANGELES, CALIFORNIA, USA, APRIL 4-11, 1987. J CELL BIOCHEM SUPPL. (1987) 0 (11 PART C), 194.
CODEN: JCBSD7.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Mathematical Biology and Statistical Methods 04500
Radiation - Radiation and Isotope Techniques 06504
Comparative Biochemistry, General 10010
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - General Biophysical Techniques 10504
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT CRYSTALLOGRAPHY PROFILE COMPARISON ENERGY CALCULATIONS

L78 ANSWER 30 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1987:1P140[illegible] BIOSIS
DN BR32:86534
TI ATOMIC SOLVATION PARAMETERS AND HYDROPHOBIC MOMENTS FOR ASSESSING PROTEIN AND LIGAND STABILITY.
AU **EISENBERG D**; YAMASHITA M; WILCOX W; TALAFOUS J; WESSON M
CS MOLECULAR BIOLOGY INST., UNIV. CALIFORNIA, LOS ANGELES, CA 90024.
SO THIRTY-FIRST ANNUAL MEETING OF THE BIOPHYSICAL SOCIETY, NEW ORLEANS, LOUISIANA, USA, FEBRUARY 22-26, 1987. BIOPHYS J. (1987) 51 (2 PART 2), 2.A.
CODEN: BIOJAU. ISSN: 0006-3495.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biochemical Studies - Minerals 10060
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT METAL ION BINDING

L78 ANSWER 31 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1987:158414 BIOSIS
DN BR33:76276
TI CYLINDRICAL BETA STRUCTURE A HYPOTHETICAL PROTEIN STRUCTURE.
AU PRYCIAK P M; CONWAY J F; EISERLING F A; **EISENBERG D**
CS MOLECULAR BIOL. INST., UNIV. CALIF., LOS ANGELES, CALIF. 90024.
SO OSHEROFF, D. L. (ED.). UCLA (UNIVERSITY OF CALIFORNIA-LOS ANGELES) **SYMPOSIA** ON MOLECULAR AND CELLULAR BIOLOGY NEW SERIES, VOL. 59. PROTEIN STRUCTURE, FOLDING, AND DESIGN; GENEX-UCLA **SYMPOSIUM**, KEYSTONE, COLORADO, USA, MARCH 30-APRIL 6, 1985. XVIII+332P. ALAN R. LISS, INC.: NEW YORK, NEW YORK, USA. ILLUS. (1986) 0 (0), 243-246.
CODEN: USMBD6. ISSN: 0735-9543. ISBN: 0-8451-2638-5.
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Virology - Bacteriophage *35504
BC Bacterial Viruses - Unspecified 02110
IT Miscellaneous Descriptors
PHAGE T4 GP48 PROTEIN

L78 ANSWER 32 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1986:446269 BIOSIS
DN BR31:102679
TI FREE ENERGY AND PROTEIN FOLDING HYDROPHOBIC MOMENTS AND SOLVATION FREE ENERGY.
AU **EISENBERG D**; WILCOX W; ESHITA S
CS MOLECULAR BIOL. INST., UNIV. CALIFORNIA, LOS ANGELES, CALIF. 90024.
SO FLETTERICK, R. AND M. ZOLLER (ED.). CURRENT COMMUNICATIONS IN MOLECULAR BIOLOGY: COMPUTER GRAPHICS AND MOLECULAR MODELING; **MEETING**, COLD SPRING HARBOR, N.Y., USA, DEC. 13-15, 1985. IX+145P. COLD SPRING HARBOR LABORATORY: COLD SPRING HARBOR, N.Y., USA. ILLUS. PAPER. (1986) 0 (0), 59-66.
ISBN: 0-87969-193-X.
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
General Biology - Information, Documentation, Retrieval and Computer Applications *00530
Biochemical Studies - Proteins, Peptides and Amino Acids *10064

Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
COMPUTER GRAPHICS

L78 ANSWER 33 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1986:232110 BIOSIS
DN BR30:114606
TI DESIGN OF SYNTHETIC FOUR-HELICAL BUNDLE PROTEIN.
AU DEGRADO W F; HO S P; WEBER P C; SALEMME F R; WILCOX W; ESHITA S; PRYCIAK P; **EISENBERG D**
CS E. I. DU PONTE DE NEMOURS AND CO., CENTRAL RES. AND DEV. DEP., EXPERIMENTAL STATION, WILMINGTON, DE 19898.
SO 30TH ANNUAL MEETING OF THE BIOPHYSICAL SOCIETY, SAN FRANCISCO, CALIF., USA, FEB. 9-13, 1986. BIOPHYS J. (1986) 49 (2 PART 2), 573A.
CODEN: BIOJAU. ISSN: 0006-3495.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Methods - Proteins, Peptides and Amino Acids *10054
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT HYDROPHOBIC INTERACTION TETRAMERIZATION

L78 ANSWER 34 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1985:144814 BIOSIS
DN BR29:34810
TI AMPHIPHILICITY AND HYDROPHOBICITY IN PROTEIN FOLDING.
AU **EISENBERG D**; WEISS R M; TERWILLIGER T C; ESHITA S
CS MOLECULAR BIOL. INST., UNIV. CALIF., LOS ANGELES, CA 90024.
SO SYMPOSIUM ON PROTEIN STRUCTURE, FOLDING AND DESIGN HELD AT THE 14TH ANNUAL UCLA (UNIVERSITY OF CALIFORNIA-LOS ANGELES) SYMPOSIA ON MOLECULAR AND CELLULAR BIOLOGY, MAR. 30-APR. 6, 1985. J CELL BIOCHEM SUPPL. (1985) 0 (9 PART B), 128.
CODEN: JCBSD7.
DT **Conference**
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
IT Miscellaneous Descriptors
ABSTRACT

L78 ANSWER 35 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1983:12287 BIOSIS
DN BR24:12287
TI THE STRUCTURE OF MELITTIN IN THE FORM I CRYSTALS AND ITS IMPLICATIONS FOR MELITTINS LYTIC AND SURFACE ACTIVITIES.
AU TERWILLIGER T C; WEISSMAN L; **EISENBERG D**
CS DEP. PHYSICS, UNIV. VA., CHARLOTTESVILLE, VA. 22901, USA.
SO **MEETING** OF THE BIOPHYSICAL SOCIETY ON PROTEIN-LIPID INTERACTIONS IN MEMBRANES, AIRLIE, VA., USA, OCT. 3-6, 1981. BIOPHYS J. (1982) 37 (1), 353-359.
CODEN: BIOJAU. ISSN: 0006-3495.
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biochemical Studies - Lipids *10066
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Membrane Phenomena *10508

Toxicology - General; Methods and Experimental *22501
Invertebrata, Comparative and Experimental Morphology, Physiology and Pathology - Insecta - Physiology *64076
BC Hymenoptera 75326
IT Miscellaneous Descriptors
HONEY BEE VENOM MEMBRANE LIPID BI LAYER INTERACTION
RN 20449-79-0Q, 37231-28-0Q (MELITTIN)
20449-79-0DQ, 37231-28-0DQ (MELITTINS)

L78 ANSWER 36 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1983:12183 BIOSIS
DN BR24:12183
TI NATURE OF MELITTIN PHOSPHO LIPID INTERACTION.
AU GHOSHIOU S; **THOMPSON M**; MUKHOPADHYAY A K
CS BIOPHYSICS CHEMICAL PHYSICS LAB., DEP. PHYSICS, UNIV. TENNESSEE, KNOXVILLE, TENNESSEE 37916.
SO **MEETING** OF THE BIOPHYSICAL SOCIETY ON PROTEIN-LIPID INTERACTIONS IN MEMBRANES, AIRLIE, VA., USA, OCT. 3-6, 1981. BIOPHYS J. (1982) 37 (1), 159-161.
CODEN: BIOJAU. ISSN: 0006-3495.
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biochemical Studies - Lipids *10066
Biophysics - General Biophysical Studies *10502
Biophysics - Molecular Properties and Macromolecules *10506
Biophysics - Membrane Phenomena *10508
Toxicology - General; Methods and Experimental *22501
Economic Entomology - Animal Pests 60012
Invertebrata, Comparative and Experimental Morphology, Physiology and Pathology - Insecta - Physiology *64076
Invertebrate Body Regions and Structures - Special Organs 64218
BC Hymenoptera 75326
IT Miscellaneous Descriptors
BEE VENOM FLUORESCENCE QUENCHING DATA

L78 ANSWER 37 OF 37 BIOSIS COPYRIGHT 2001 BIOSIS
AN 1981:53774 BIOSIS
DN BR20:53774
TI STRUCTURAL STUDIES OF BEE MELITTIN.
AU **EISENBERG D**; TERWILLIGER T C; TSUI F
CS MOL. BIOL. INST., UNIV. CALIF., LOS ANGELES, CALIF. 90024, USA.
SO **MEETING** ON PROTEINS AND NUCLEOPROTEINS: STRUCTURE, DYNAMICS AND **ASSEMBLY**, AIRLIE, VA., USA, MAY 18-21, 1980. BIOPHYS J. (1980) 32 (1), 252-254.
CODEN: BIOJAU. ISSN: 0006-3495.
FS BR; OLD
LA English
CC **General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals 00520**
Biochemical Studies - Proteins, Peptides and Amino Acids *10064
Biophysics - Molecular Properties and Macromolecules *10506
Toxicology - General; Methods and Experimental *22501
Invertebrata, Comparative and Experimental Morphology, Physiology and Pathology - Insecta - Physiology *64076
BC Hymenoptera 75326
IT Miscellaneous Descriptors
VENOM CONSTITUENT
RN 20449-79-0Q, 37231-28-0Q (MELITTIN)

=> fil wpix

FILE 'WPIX' ENTERED AT 13:32:46 ON 12 FEB 2001

COPYRIGHT (C) 2001 DERWENT INFORMATION LTD

FILE LAST UPDATED: 07 FEB 2001 <20010207/UP>
>>>UPDATE WEEKS:
MOST RECENT DERWENT WEEK 200108 <200108/DW>
DERWENT WEEK FOR CHEMICAL CODING: 200108
DERWENT WEEK FOR POLYMER INDEXING: 200108
DERWENT WORLD PATENTS INDEX SUBSCRIBER FILE, COVERS 1963 TO DATE

>>> FOR UP-TO-DATE INFORMATION ABOUT THE DERWENT CHEMISTRY RESOURCE, PLEASE VISIT http://www.derwent.com/chemistryresource/index.html <<<

>>> FOR DETAILS OF THE PATENTS COVERED IN CURRENT UPDATES, SEE http://www.derwent.com/dwpcodes.html <<<

=>

=>

=> d all abeq tech tot 194

L94 ANSWER 1 OF 3 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
AN 2000-686550 [67] WPIX
DNN N2000-507601 DNC C2000-208765
TI Functionally linked **polypeptide** identification comprising outputting one **protein** indicating functional link for alignment between primary amino acid sequences of distinct non-homologous **polypeptides**.
DC B04 D16 T01
IN **EISENBERG, D**; GROTHE, R; **MARCOTTE, E**; **PELLEGRINI, M**; **THOMPSON, M**; **YEATES, T**
PA (REGC) UNIV CALIFORNIA
CYC 90
PI WO 2000045322 A1 20000803 (200067)* EN 74p G06F019-00 <--
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CR CU CZ DE DK DM EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
AU 2000027442 A 20000818 (200067) G06F019-00 <--
ADT WO 2000045322 A1 WO 2000-US2246 20000128; AU 2000027442 A AU 2000-27442 20000128
FDT AU 2000027442 A Based on WO 200045322
PRAI US 1999-134093 19990514; US 1999-117644 19990129; US 1999-118206 19990201; US 1999-126593 19990326; US 1999-134093 19990514
IC ICM **G06F019-00**
AB WO 200045322 A UPAB: 20001023
NOVELTY - Identifying multiple **polypeptides** as functionally-linked comprising aligning a primary amino acid (aa) sequence of multiple distinct non-homologous **polypeptides** to the primary aa sequence of a number of **proteins** and, for any alignment found, outputting an indication identifying the **protein** as an indication of a functional link between **polypeptides**, is new.
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:
(a) computer program stored on a computer readable medium for identifying the **protein** as functionally linked;
(b) a method for determining evolutionary distance of two **proteins**;
(c) a computer program stored on computer readable medium, for

determining evolutionary distance between two **proteins**; and
(d) a method for determining functional links between two **polypeptides**.
USE - The method is useful for determining **protein** functions from genomic sequences for use in medical and agricultural biotechnology for identifying new genes and identifying potential targets from pharmaceutical compounds.
ADVANTAGE - The method enables identification of the complex or pathway that a **protein** participates in, by comparing profiles and also reduces error in functional links.
DESCRIPTION OF DRAWING(S) - The figure shows a flow diagram describing the Rosetta Stone method of identifying functionally linked **polypeptide**.
Dwg.2A/12

FS CPI EPI
FA AB; GI; DCN
MC CPI: B11-C08E; B12-K04; D05-H09
EPI: **T01-J04A**; T01-S03
TECH UPTX: 20001223
TECHNOLOGY FOCUS - BIOTECHNOLOGY - Preferred Method: The aligning is performed by an algorithm selected from Smith-Waterman algorithms, Needleman-Wunsch algorithm, BLAST (basic local alignment search tool), FASTA (not defined) and PSI-BLAST. The multiple distinct non-homologous **polypeptides** are from a database and the multiple distinct non-homologous **polypeptides** are obtained by translating a nucleic acid sequence from a genome database. Alignment is based on the degree of homology of the multiple distinct non-homologous **polypeptides** to the **proteins**. The **protein** contains a fragment of the primary amino acid sequence of each of the multiple distinct non-homologous **polypeptides**. The method further comprises determining the significance of the **protein** by computing a probability (p) value threshold. The probability threshold is set with respect to the value 1/NM, based on the total number of sequence comparisons that are to be performed, where N is the number of **proteins** in a first organisms genome and M in all other genomes. For any alignment found between the primary amino acid sequence of the distinct non-homologous **polypeptides** and at least 1 primary amino acid sequence of the number of **proteins**, filtering excessive functional links between one distinct non-homologous **polypeptide** and an excessive number of other distinct non-homologous **polypeptides**.

L94 ANSWER 2 OF 3 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
AN 1999-081222 [07] WPIX
DNN N1999-058380 DNC C1999-024468
TI Use of bactericidal/permeability-increasing **protein** - to form crystals to provide a 3-dimensional modelling to provide rational drug design for mimetics and ligands for BPI and related **proteins**.
DC B04 C03 C07 J04 T01
IN BEAMER, L J; CARROLL, S F; ELSENBERG, D; **EISENBERG, D**
PA (REGC) UNIV CALIFORNIA; (XOMA) XOMA CORP; (XOMA) XOMA
CYC 83
PI WO 9858961 A1 19981230 (199907)* EN 203p C07K014-47 <--
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA PT SD SE SZ UG ZW
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG US UZ VN YU ZW
AU 9884722 A 19990104 (199921) C07K014-47 <--
EP 1007557 A1 20000614 (200033) EN C07K014-47 <--
R: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT MC NL PT SE
US 6093573 A 20000725 (200038) G01N033-48
ADT WO 9858961 A1 WO 1998-US13007 19980622; AU 9884722 A AU 1998-84722 19980622; EP 1007557 A1 EP 1998-930480 19980622, WO 1998-US13007 19980622; US 6093573 A US 1997-879565 19970620

FDT AU 9884722 A Based on WO 9858961; EP 1007557 A1 Based on WO 9858961
PRAI US 1997-879565 19970620
IC ICM **C07K014-47**; G01N033-48
ICS **C07K001-00; G06F017-50**
AB WO 9858961 A UPAB: 19990224
Use of atomic coordinates of bactericidal/permeability-increasing (BPI) **protein**, or fragment analogue or variant, to model a BPI **protein**. Also claimed are: (1) the use of atomic coordinates of BPI **protein**, or fragment, analogue or variant, to model a BPI-related lipid transfer **protein**; (2) the use of atomic coordinates of BPI **protein** to computationally design a chemical compound for mimicking BPI **protein**, or fragment, analogue or variant; (3) the use of atomic coordinates of BPI **protein** to computationally design a chemical compound for mimicking a BPI-related lipid transfer **protein**, or fragment, analogue or variant; (4) the use of atomic coordinates of BPI **protein**, to design a chemical compound capable of associating with a BPI-related lipid binding **protein**, or fragment, analogue or variant; (5) the use of atomic coordinates of BPI **protein** to design a model of ligands in an active site of a lipid binding **protein**; (6) the use of atomic coordinates of BPI to design compounds with at least one activity selected from antibacterial, antifungal, antimycobacterial, antichlamydial, antiprotozoan, heparin-binding, endotoxin-binding, heparin-neutralising, endotoxin-neutralising, inhibition of tumour and endothelial cell proliferation, inhibition of angiogenesis, anti-inflammatory, anticoagulant and antithrombolytic; (7) a method of 3-dimensional (3D) modelling of a BPI **protein** comprising: (a) providing 3D atomic coordinates derived from X-ray diffraction measurements of a BPI **protein** in a computer readable format; (b) inputting the data from (a) into a computer with appropriate software programmes; (c) generating a 3D structural representation of the BPI **protein** suitable for visualisation and further computational manipulation; (8) a method of 3D modelling of a BPI-related lipid transfer **protein** comprising: (a) providing 3D atomic coordinates derived from X-ray diffraction measurements of a BPI **protein** in a computer readable format; (b) inputting the data from (a) into a computer with appropriate software programmes; (c) generating a 3D structural representation of the BPI-related lipid transfer **protein** suitable for visualisation and further computational manipulation; (9) a method for providing an atomic model of a BPI **protein** or fragment, analogue or variant, comprising: (a) providing a computer readable medium (CRM) having stored on it atomic coordinate/x-ray diffraction data of the BPI **protein**, or fragment, analogue or variant, in crystalline form, the data sufficient to model the 3D structure of the BPI **protein** or fragment, analogue or variant; (b) analysing on a computer, using at least one subroutine executed in the computer, atomic coordinate/x-ray diffraction data from (a) to provide atomic coordinate data output defining an atomic model of the BPI **protein**, or fragment, analogue or variant, the analysing utilising at least one computing algorithm selected from data processing and reduction, auto-indexing, intensity scaling, intensity merging, amplitude conversion, truncation, molecular replacement, molecular alignment, molecular refinement, electron density map calculation, electron density modification, electron map visualisation, model building, rigid body refinement, positional refinement; and (c) obtaining atomic coordinate data defining the 3D structure of at least one of the BPI **protein**, or fragment, analogue or variant; (10) a computer-based system for providing atomic model data of the 3D structure of BPI **protein**, or fragment, analogue or variant, a BPI mutant or a BPI fragment, comprising the following elements: (a) at least one CRM having stored on it atomic coordinate/x-ray diffraction data of the BPI **protein**, or fragment, analogue or variant; (b) at least one computing subroutine that, when executed in a computer, causes the computer to analyse atomic coordinate/x-ray diffraction data from (a) to provide atomic coordinate data output defining an atomic model of the BPI **protein**, or fragment, analogue or variant, the analysing utilizing at least one

computing subroutine selected from: data processing and reduction, auto-indexing, intensity scaling, intensity merging, amplitude conversion, truncation, molecular replacement, molecular alignment, molecular refinement, electron density map calculation, electron density modification, electron map visualisation, model building, rigid body refinement, positional refinement; and (c) a retrieval device for obtaining atomic coordinate output data substantially defining the 3D structure of the BPI **protein**, or fragment, analogue or variant; (11) a method for providing a computer atomic model of a ligand of a BPI **protein** or fragment, analogue or variant, comprising: (a) providing a CRM having stored on it atomic coordinate data of a BPI **protein**, or fragment, analogue, or variant; (b) providing a CRM having stored on it atomic coordinate data to generate atomic models of potential ligands of the BPI **protein**, or fragment, analogue, or variant; (c) analysing on a computer, using at least one subroutine executed in the computer, the atomic coordinate data from (a) and ligand data from (b), to determine binding sites of BPI **protein**, or fragment, analogue or variant, and to provide atomic coordinate data defining an atomic model of at least one ligand of the BPI, BPI mutant or a fragment, the analysing utilising computing subroutines selected from the group consisting of data processing and reduction, auto-indexing, intensity scaling, intensity merging, amplitude conversion, truncation, molecular replacement, molecular alignment, molecular refinement, electron density map calculation, electron density modification, electron map visualisation, model building, rigid body refinement, positional refinement; and (d) obtaining atomic coordinate model output data defining the 3D structure of the at least one ligand of the BPI **protein**, or fragment, analog, or variant, and (12) a computer-based system for providing an atomic model of at least one ligand of a BPI, BPI mutant or a fragment, comprising the following elements; (a) a CRM having stored on it atomic coordinate data of a BPI, mutant or fragment; (b) a CRM having stored on it atomic coordinate data to generate atomic models of potential ligands of a BPI, mutant or fragment; (c) at least one computing subroutine for analysing on a computer, the atomic coordinate data from (a) and (b), to determine binding sites of BPI **protein**, or fragment, analogue, or variant, and to provide data output defining an atomic model of at least one potential ligand of BPI **protein**, or fragment, analogue or variant, the analysing utilising at least one computing subroutine selected from data processing and reduction, auto-indexing, intensity scaling, intensity merging, amplitude conversion, truncation, molecular replacement, molecular alignment, molecular refinement, electron density map calculation, electron density modification, electron map visualisation, model building, rigid body refinement, positional refinement; and (d) a retrieval device for obtaining atomic coordinate data of the at least one ligand of a BPI **protein** or fragment, analogue or variant.

USE - The methods can be used for molecule modelling of BPI and related **proteins** and rational drug design of mimetics and ligands for BPI and for related **proteins**. They can also be used to provide mutants of BPI or fragments, analogues or variants characterised by one or more different properties as compared with wild-type BPI. These properties include altered surface charge, altered lipid binding pockets, altered specificity or higher activity. They can be used to design compounds with at least one activity selected from antibacterial, antifungal, antimycobacterial, antichlamydial, antiprotozoan, heparin-binding, endotoxin-binding, heparin-neutralising, endotoxin-neutralising, inhibition of tumour and endothelial cell proliferation, inhibition of angiogenesis, anti-inflammatory, anticoagulant and antithrombolytic.

Dwg.0/6

FS CPI EPI

FA AB

MC CPI: B04-L04; C04-L04; B04-N02; C04-N02; B11-C08; C11-C08; B12-K04E; C12-K04E; J04-B01

EPI: **T01-J10C4; T01-J15H**

L94 ANSWER 3 OF 3 WPIX COPYRIGHT 2001 DERWENT INFORMATION LTD
AN 1993-045645 [05] WPIX
DNN N1993-034929 DNC C1993-020663
TI Characterising the three-dimensional structure of a **protein** - by analysing aminoacid residue positions and comparing with known **protein** structures.
DC B04 D16 T01
IN BOWIE, J U; **EISENBERG, D**; LUTHY, R
PA (REGC) UNIV CALIFORNIA
CYC 18
PI WO 9301484 A1 19930121 (199305)* EN 56p G06F015-20 <--
RW: AT BE CH DE DK ES FR GB GR IT LU MC NL SE
W: AU CA JP
AU 9224082 A 19930211 (199321) G06F015-20 <--
US 5436850 A 19950725 (199535) 23p G06F019-00 <--
ADT WO 9301484 A1 WO 1992-US5773 19920710; AU 9224082 A AU 1992-24082 19920710; US 5436850 A Cont of US 1991-728640 19910711, US 1994-218685 19940328
FDT AU 9224082 A Based on WO 9301484
PRAI US 1991-728640 19910711; US 1994-218685 19940328
REP US 4704692; US 4717653; US 4853871; US 4881175; US 4908773; US 4939666; US 4946778; US 4976958; US 5087558
IC ICM **G06F015-20; G06F019-00**
ICS C12N015-00; C12Q001-68
AB WO 9301484 A UPAB: 19931119
Characterising the 3-dimensional(3-D) structure of a **protein**, comprises (a) determining, from the 3-D structure of the **protein**, values for a structural properties P1,P2....Pn for each amino acid residue position of the **protein**, (b) assigning each residue of the **protein** to one environment class based upon the values for the n structural properties P1,P2 ....Pn for the residue, thereby generating a 1-dimensional environment string comprising the environment class of each residue in the 3-D **protein** structure.
USE/ADVANTAGE - Permit the assignment of many amino acid sequences to known 3-D structures. Used partic. for screening structural analogues of a known **protein** sequence. The 3-D compatibility searches are able to detect structural relationships that may not be apparent by sequence similarity.
1/5
Dwg.1/5
FS CPI EPI
FA AB; GI
MC CPI: **B04-B04A**; B12-K04; D05-H09
EPI: **T01-J10B2**
ABEQ US 5436850 A UPAB: 19950905
The three-dimensional structure of a **protein** is characterised by determining values for n structural properties P1-Pn for each amino acid residue, and assigning each residue to one of a number of environmental classes based on the values to generate a one-dimensional environment string comprising the class of each residue. The data generated are input into a programmed computer which compares them to a database of other **proteins** of known structure and outputs analogous structures. The properties pref. include the total area of a residue side-chain buried by other **protein** atoms inaccessible to solvent, the fraction of the side-chain area covered by polar atoms or water, and the local secondary structure.
USE/ADVANTAGE - Partic. for identifying **protein** sequences which fold into a known three-dimensional structure. Relates a one-dimensional target sequence directly to known three-dimensional structures and effectively utilises information about the accommodation of sequence changes inherent in known structures.
Dwg.1/8